TOWARDS OPTIMAL RHYTHM

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF LINGUISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Stephanie Sin-yun Shih
March 2014

This dissertation is online at: http://purl.stanford.edu/rr762kf8691

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Arto Anttila, Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Sharon Inkelas, Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Joan Bresnan

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Vera Gribanov

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

This thesis argues that rhythmic well-formedness preferences contribute to conditioning morphosyntactic choices, providing evidence from patterns in language use that constraints on phonological constructs are at work in the assessment of competing morphosyntactic variants. The results of the thesis call into question a fundamental empirical assumption underlying many standard models of grammar and of language production: that metrical or segmental phonology cannot influence morphosyntactic encoding.

Phonologically-conditioned morphological phenomena are of familiar stock. Phonological constraints can force blocking of morphological processes and combinatorics, resulting in a number of repair strategies: re-ordering, periphrasis, deletion, and suppletion. It is shown in this thesis that phonologically-conditioned syntactic phenomena follow a similar typological spread. The same phonological constraints that interfere with morphology also operate across word and phrase boundaries, triggering repairs of word (re-)ordering, periphrasis/paraphrasing, deletion, and suppletion (i.e., lexeme replacement).

Two empirical studies of the rhythmic conditioning of word choice (e.g., personal name choice) and syntactic choice (e.g., genitive alternation) in English are presented. The case studies demonstrate that rhythmic optimization, in addition to other phonological well-formedness preferences such as phonotactic co-occurrence restrictions, are active in word and construction variation in syntagmatic contexts. It is furthermore shown that the effect of rhythm is closely tied to semantic factors such as animacy, which reveals that rhythm must interact and compete with non-phonological constraints in the system.

Allowing interaction between phonological material and morphosyntactic choices raises the issue of how much surface and underlying phonetic and phonological information is available at the point of morphosyntactic computation. Rhythm offers a natural test case of the availability of underlying versus post-lexically specified information via the distinction in stress properties of lexical (content) and grammatical (function) words. A large-scale corpus study of content and function word stress in conversational American English is presented. Results of the study point to complex differences between word categories in terms of underlying and surface stress properties. These differences in stress not only trigger differences in rhythmic optimization by word category but they also demonstrate that morphosyntactic competitors are assessed without consideration of the potential output of surface rhythmic optimization. In contrast, evidence from end weight phenomena suggests that lexically-encoded information about underlying phonological stress is available during morphosyntactic computation.

The view that emerges from the empirical studies in this thesis is one that allows for potential phonological influence on morphological and syntactic outputs. The phonological constraints that are most active will necessarily be ones that regulate syntagmatic effects that occur when words combine in the morphosyntax, and these phonological constraints—including the propensity towards optimal rhythm—must compete for satisfaction against other active, non-phonological pressures.

Acknowledgements

This dissertation is one about interfaces, and so the acknowledgements are no different. I have been incredibly fortunate to spend my formative linguist years at the confluence of two powerhouse universities and with the many linguists here in the Greater Bay Area. As I've discovered, working at the interface is never easy, but it is far more rewarding than any one component could ever be on its own. In this undertaking, I am thankful to many for their help, guidance, time, patience, and support along the way.

Although he did not have a direct hand in this work, I am first indebted to Paul Kiparsky, without whom none of my linguistics career would have happened. He inspired and advised those who in turn inspired and advised me, and he was the one who brought me to Stanford, in a move that started me down a path that neither he nor I had ever imagined I would take.

Acting as a guide down this path has been the ever-calm Sharon Inkelas, whose remarkable skill for rescuing confused ideas and lost people (as well as confused people and lost ideas) has benefited me innumerable times over the years. Sharon was the first to incubate my early, undergraduate interest in suprasegmental phonology. Since then, even over long temporal and geographical distances, she has helped in critical times to see through my developing interest in suprasegmental phonology at the interface. Through the years, Sharon has somehow stubbornly believed in my potential as a linguist even when empirical data supporting such a belief has been severely lacking, and I am deeply grateful to have such a lofty potential (and challenge) to strive towards. I must also thank Sharon in her former capacity as Berkeley Linguistics chair for arranging a concrete motivation for me to finish the dissertation.

One of the most valuable lessons of science that I took to heart early on during my time at Stanford is to doggedly pursue the question, "What does this *predict*?" I must credit Arto Anttila for this lesson, which has been pivotal in sharpening my approach to theory. Through the hours of discussions in his office, I like to think that I

have absorbed at least some of Arto's admirable tenacity for following thoughts through to their (however bitter) logical ends.

Joan Bresnan deserves a great deal of credit for taking in an orphan phonologist in 2008 and kindly letting me hijack the genitives project during its inaugural summer to examine rhythmic effects. Since then, she has graciously shared with me any and all data that I dare to ask for; provided me with funding for many summers of work; and pointed me to the tools that I need to answer the questions that I sometimes don't even know I will ask yet. From working with Joan, I have absorbed a deep appreciation of quantitative and empirical investigation as well as a certain fastidiousness with numbers: any proper statistics in this dissertation are thanks to her keen influence. Joan also never lets me forget that hard work and having *fun* (while working hard) go hand-in-hand.

Vera Gribanova has been an inspiring role model for me in learning how to be an early-career academic and a linguist in general. Without her lead, I would most surely be a lesser scholar today. Vera has had so much open-mindedness and patience in discussing with me fairly radical ideas of the phonology-morphosyntax interface and of grammar, and these discussions have greatly improved and tempered my own thoughts on the subject, even if Vera herself does not agree with them. I should also officially apologize to her for dragging her into the wonderful ordeal that is our Morphosyntax-Phonology Connection workshop and book project, but I must say that in the future, I hope we get the chance to do it again!

Although I was wisely advised to stop at four members on my official dissertation committee, many more faculty deserve thanks in aiding me with this endeavor. At Stanford, without Beth Levin's tireless and proactive guidance, I would have never managed to navigate or make it through graduate school. Beth also taught me how to make pecan pie and appreciate plums, which are essential life lessons no matter where I end up. One of my biggest regrets of graduate school was not working more extensively with Dan Jurafsky. As a first step towards remedying this mistake, one of the wisest choices I made during graduate school was having Dan on the thesis defense committee. I must thank him for his energetic enthusiasm and insights at the eleventh hour. Meghan Sumner was the external member at my dissertation proposal meeting and has, over the years, contributed influential perspectives to my thinking about processing and variation. I am so thankful that Rob Podesva made the move to Stanford towards the end of my time here, and I only wish it'd been sooner. While Rob has imparted much valuable advice about linguistics as a field, I am even more appreciative of his support in the form of friendship, lots of delicious (homemade) macarons, and many hours—together with Sharon—of very necessary musical diversions. Thank you also to the staff of Stanford Linguistics for all of their help during my time.

At Berkeley, Kristin Hanson is to thank for my start in linguistics, and for making it her personal mission through the years to ensure that this dissertation and my graduate studies were, at some level, "a form of self-expression and passion." Keith Johnson, Larry Hyman, and Susanne Gahl also contributed discussion to various portions of this (or other phonological) work.

Many thanks are due to Kie Zuraw and the linguistics community at UCLA for providing an unofficial intellectual home in the final stages of dissertation-writing and the first stages of post-dissertation research. Collaborating with Kie on the next steps arising out of my current work has served to sharpen the overall exposition of my research program. I should also thank Bruce Hayes for his curiously prophetic nuggets of advice on my career trajectory in linguistics that he imparts onto me every few years.

Throughout graduate school, I have somehow managed to surround myself with the most incredible group of peers and friends a person could ask for. In every way, they inspire me to be even better just to be worthy of keeping up. Lauren Hall-Lew, despite our eight hour time difference, has been my daily work partner, cheerleader, advice-giver, and answer-all-of-Stephanie's-ridiculous-questions-about-sociolinguistics-er. Rebecca Starr is in the unique position of having read and commented on nearly every piece of writing that I have produced in the latter half of my graduate career: every academic needs a friend like this. Jason Grafmiller has endless patience in having me as an oft-collaborator and quantitative methods partner-in-crime, and I hope his patience never runs out. For general graduate school camaraderie, thanks to

the fellow students at Stanford, Berkeley, and beyond, and especially to my 七福神 Stanford cohort (Kyuwon Moon, SK Kim, Sven Lauer, Rob Munro), my academic half-brother Matt Adams, Clara Cohen, Laura McPherson, Nathan Schneider, and Kashmiri Stec. Special acknowledgement should be given to Mike Speriosu, who was such an efficient programming teacher that it took him only one afternoon during my first year (traded for a home-cooked dinner, naturally) to show me the Python ropes.

Beyond linguistics, Jessica Chow has been an unwavering friend and fellow workaholic since we first found each other in fifth grade honor band. Lynn Wang has been there since we first despised each other in first grade (thank goodness we grew out of that). Thank you to Andréa Davis, for her ever-growing friendship since our first days together in Sharon Inkelas's undergraduate phonology class. In addition to suffering me as a roommate for many years, María Helga Guðmundsdóttir instilled in me an appreciation for the subtle complexities of Brahms, thus providing me with a soundtrack for my dissertation years. Toni Bird/Gauthier always encouraged my creative outlets whenever my sanity needed it the most (and Milo and Mogul always provided simple canine love when that was needed the most). Many thanks to Joy Kao and Minh Thai for generously hosting me (and providing Porto's) whenever I was visiting Los Angeles. Lastly, to Richard, who taught me over α years together that relying on someone can actually make you stronger. Richard also deserves credit for having enough immense patience to discuss mathematics with me and be my personal IT guy, for which I should probably owe him a lifetime of boba tea drinks.

Gratitude is long overdue to my family, who, in the infinitely paradoxical way of families, pushed me hard to get into graduate school and, once I'd arrived, constantly prodded me not to work too hard. Every day I am thankful for how much my parents have meticulously planned and saved in order to provide me roofs under which I could carry out my studies for the many, many years I have been in school. Such a debt I will never be able to repay, but I hope that being able to boast that their daughter has a doctorate from Stanford will bring my parents at least a small bit of redeeming joy.

Finally, this dissertation is dedicated to the teachers from my most formative years, pre-linguistics: Mrs. Waid, Mr. Crawford, Ms. Connaughton, Ms. Young, Ms. Nitta, Ms. Storer—and especially to Mrs. Kollewee and Mrs. MacAuley, who together decided that I needed to learn how to write, and to Ms. Cone, who decided that I needed to learn how to read more carefully. This dissertation is not beautifully written nor does it reflect someone terribly well-read, but I hope that it still does all of them proud.

This material is based in part upon work supported by the National Science Foundation under Grant Number BCS-1025602 to Stanford University for the research project "The Dynamics of Probabilistic Grammar" (PI Joan Bresnan). This work was also supported in part by the Mabelle McLeod Lewis Memorial Fund. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author's and do not necessarily reflect the views of the National Science Foundation or the Mabelle McLeod Lewis Memorial Fund.

Preface

When shall syntax, phonetics, and phonology meet again?
In processing, production, or in perception?

When stress clash is done,
When optimal rhythm has been lost and won.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

"Taste, like rhythm, may be described,
but it does not exist until it is experienced."
– *Marcella Hazan*

1 Introduction

A cornerstone of modern phonological study is the observation that multiple, heterogeneous "repairs" conspire to optimize phonological structure in response to a given phonological well-formedness condition (Kisseberth 1970; Prince & Smolensky 1993; a.o.). For example, the avoidance of nasal consonant + voiceless consonant sequences (e.g., *NC̥) engenders a number of diverse phonological repair strategies cross-linguistically, including deletion, fusion, assimilation, and dissimilation (e.g., Pater 1999). Similarly, rhythmic well-formedness conditions (i.e., avoiding adjacent stressed syllables or adjacent unstressed syllables) trigger a number of repairs ranging from stress retraction or reduction to vowel lengthening or syncope (e.g., Nespor & Vogel 1989; Gouskova 2003; McCarthy 2008). Typically, such repairs take the form of phonological (or phonetic) processes that can optimize the output of phonological conditions. There are, however, extra-phonological possibilities as well. Amongst the most commonly noted are morphological processes that "repair" or avoid ill-formed phonological structure: suppletive allomorphy, periphrasis, variable affixation or affix reordering, and blocking (Poser 1992; McCarthy & Prince 1993a; McCarthy & Prince 1993a; Prince & Smolensky 1993; Raffelsiefen 2004; McCarthy & Wolf 2005; Paster 2006; a.o.).

This dissertation argues that the optimization of phonological well-formedness is not affected solely by word-internal morphological, phonological, or phonetic repair processes and that phonological considerations also underlie alternations in the syntactic domain. The simple claim that phonology can interfere with syntax is not itself new: for example, there is a sizable literature claiming that Heavy Noun Phrase Shift, in which heavier constituents (e.g., ones with more prosodic phrase constituents or a greater number of stressed syllables) are licensed in specific syntactic positions, is phonologically conditioned (Zec & Inkelas 1990) and even potentially phonologically optimizing (Zubizarreta 1998; Anttila et al. 2010; though cf. Hawkins 1994; Gibson 2000; Wasow 2002 for alternative analyses that are not prosodically-driven).[1] This dissertation looks beyond Heavy NP shift to other syntactic phenomena, focusing in particular on those syntactic alternations in English which optimize phonological rhythm. The reasons for choosing to examine rhythm is that, because of the nature of English metrical structure, every sentence provides a test case for rhythmic optimization; therefore, rhythm is a logical place to start to probe for phonological influences on syntax. This dissertation draws from morphosyntactic variation in natural language use to show that lexical, word order, and construction choices contribute to optimizing phonological structure at the metrical level (cf. the prosodic hierarchy[2]) (see also Schlüter 2005; a.o.). This study also addresses the prioritization of potential repairs for syntactic constructions that violate phonological well-formedness, and shows that morphosyntactic repairs are, in some cases, the preferred option, even when phonological and phonetic repairs are also available. The fact that morphosyntactic choice defers to phonological considerations raises implications for formal and psycholinguistic

[1] See also prosodically-driven clitic placement (e.g., Schutze 1994).

[2] I follow here a standard division of labor between the prosodic hierarchy and the metrical hierarchy (Selkirk 1986; Inkelas 1990:38; a.o.). The former contains prosodic constituents at or above the prosodic word (cf. e.g. Inkelas 1990, for sub-prosodic word constituency). The latter, in contrast, contains lower-level skeletal timing units of foot, syllable, and mora. A number of asymmetries have led to such a split in representation including, for example, recursive and labeled structure in prosodic domains versus simpler structure in metrical domains (e.g., Itô & Mester 2009).

models of the morphosyntax-phonology interface that rely on strict ordering between syntactic and phonological encoding.

1.1 Towards optimal rhythm

Rhythm—the temporally-regular recurrence of a given pattern—is a basic property of human cognitive and physiological systems. Rhythmic regularity is closely tied to cognitive, neural, and motor functions in the basal ganglia and the cerebral cortex (e.g., Grahn & Rowe 2013) and is present in heart beats, circadian habits, and bipedal movement. In addition to physical connections, rhythm pervades human expressive forms, in music, art, dance, and poetry, across domains and across cultures.

The linguistic embodiment of this basic cognitive property is the Principle of Rhythmic Alternation, which arises from the expectation that events in the speech stream will recur in regular succession (Abercrombie 1967:96). As a fundamental hypothesis underpinning modern metrical stress theory, the Principle of Rhythmic Alternation states that languages are rhythmically organized, with a propensity for the regular recurrence of strong and weak elements, as in a sequence of stressed versus unstressed syllables (Sweet 1876:12; Liberman 1975; Selkirk 1984; a.o.):

(1) supercalifragilisticexpialidocious

	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ
	sú	pɚ	kǽ	lɨ	fɹǽ	ʤə	lí	stɪk	ɛ́k	spi	ǽ	lɨ	dóʊ	ʃəs

Under the Principle of Rhythmic Alternation, the hypothesis is that languages strive towards upholding rhythmic alternation. One strong reading of the Principle of Rhythmic Alternation is that binary alternation is preferred, as in *sùpercàlifràgilìsticèxpiàlidócious* (1), where exactly one unstressed syllable occurs between stressed syllables. Though less common, regular ternary patterns are also possible[3] (Hayes 1995; Elenbaas & Kager 1999; a.o.), but quaternary patterns and rhythmic patterns that space beats farther apart decompose into smaller binary or ternary units (Selkirk 1984:37; Hayes 1984; a.o.).

The idealized, target system under the Principle of Rhythmic Alternation is thus one that promotes rhythmic regularity between syllables or stresses. A phonetic realization of such an idealized system is isochrony, defined as equal length in phonetically-measurable durations of interstress intervals, syllables, or moras (Pike 1945; Abercrombie 1967). True isochrony as a systematic, measurable realization of rhythm, however, has proved elusive in research. Studies that measure duration have not established sufficient regularity, nor have they found significant measurable differences between the perceived types of isochrony cross-linguistically[4] (Lehiste 1977; Dauer 1983; Beckman 1992; Arvaniti 2012; amongst numerous others). Extending the idea of phonetically-measurable rhythm, some have sought to redefine measurements of isochrony by incorporating phonetic information beyond duration, including differences in vocalic intervals, consonant durations, or acoustic content (Ramus et al. 1999; Grabe & Low 2002; Tilsen & Johnson 2008; Tilsen & Arvaniti 2013; a.o.); these studies have been met with varying degrees of success at uncovering isochrony. Beyond phonetics, rhythm can be considered an abstract phonological construct realized only in part by—or perceptually informed from—phonetic isochrony on the surface (Arvaniti 2009). This distinction between phonetic timing and phonological rhythm is one that has been long assumed in defining linguistic rhythm (Liberman 1975; Liberman & Prince 1977; a.o.), following the idea in music cognition research wherein perceived regularity in meter persists despite surface fluctuations in tempo (e.g., Lerdahl & Jackendoff 1983).

---

[3] Cayuvava (spoken in Bolivia) is an example of a ternary stress system: e.g., *šákahe* ‘stomach’, *čàadiròboβurúruče* ‘ninety-nine (first digit)) (Hayes 1995:309; Elenbaas & Kager 1999).

[4] Isochrony was originally proposed as a way to characterize impressionistic differences between “stress-timed” (e.g., English) ad “syllable-timed” (e.g., French) languages—also referred to as “Morse code” versus “machine gun”-sounding languages, respectively.

The Principle of Rhythmic Alternation is argued to be so pervasive in stress-based languages that it applies across morpheme and word boundaries as they combine in syntagmatic contexts (e.g., Liberman & Prince 1977; Selkirk 1984; Kaisse 1985; Nespor & Vogel 1989; Kager 1996, amongst numerous others). For example, Dutch plural suppletive allomorphy between plural allomorphs –*en* and –*s* is apparently conditioned by lapse avoidance (Paster 2006:114; citing Booij 1997):

(2)

	Singular	Plural	Gloss
a.	dám	dámm-**en**	'dam, dams'
	kanón	kanónn-**en**	'gun, guns'
	kanáal	kanáal-**en**	'channel, channels'
	lèdikánt	lèdikánt-**en**	'bed, beds'
	ólifànt	ólifànt-**en**	'elephant, elephants'
b.	kánon	kánon-**s**	'canon, canons'
	bézəm	bézəm-**s**	'sweep, sweeps'
	tóga	tóga-**s**	'gown, gowns'
	proféssor	proféssor-**s**	'professor, professors'

As shown in (2), the –*en* allomorph attaches to words with final stress (2a) while the –*s* allomorph attaches to words with penultimate stress (2b). This pattern avoids lapse that would otherwise be incurred if the stressless –*en* suffix attached to words with penultimate stress (e.g., *kánon-en). Across word boundaries, Temperley (2009) demonstrates that the overall distributional rhythmic regularity, as calculated by the number of unstressed syllables between stressed syllables, across two corpora of American English speech is significantly higher than would be expected by chance. This behavior—rhythmic optimization in syntagmatic contexts across words and how optimal rhythm in these contexts is satisfied—is the main topic of this thesis.

With the Principle of Rhythmic Alternation as a guiding hypothesis of the metrical organization of language, the critical question is how the optimization of rhythm occurs. That is, what are the strategies for dealing with non-optimal rhythm patterns that arise? "Compensation strategies"—so-called by Schlüter (2005)—that have been proposed for achieving optimal rhythmic patterning fall into three categories by domain of repair:[5]

1. phonetic adjustment,
2. phonological adjustment, and
3. morphosyntactic adjustment.[6]

Phonetic adjustments involve reduction or enhancement of the phonetic signal—for example, lengthening or shortening in the temporal domain of the speech stream—to repair rhythmic irregularities. Although true durational isochrony between stressed syllables in natural speech is rare (Lehiste 1977; Dauer 1983; Beckman 1992; Arvaniti 2012; a.o.), Tilsen et al. (2012) find in an acoustic analysis that English speakers tend to impose inter-stress regularity through micro-adjustments of delaying syllable onsets in an effort to achieve optimal rhythm.

Phonological adjustments include repairs utilizing phonological material: for example, stress shift, vowel reduction, and syncope. A particularly well-known phenomenon of phonological optimization observed in English, as well as in Italian and other languages, is the Rhythm Rule, which repairs irregular rhythm in stress clash environments across words by shifting stress to an earlier syllable (Nespor & Vogel 1989; a.o.):[7,8]

(3) eigh**téen** + **cán**dles → **éigh**teen **cán**dles

[5] Schlüter (2005) refers to these categories of compensation strategies as “phonetic”, “phonological”, and “extra-phonological” adjustments; however, because phonetics is also extra-phonological, I take Schlüter’s “extra-phonological” here to mean “morphosyntactic”—that is, of the morphological and syntactic components of grammar.

[6] Using “morphosyntax” as a shorthand for simply the union between morphology and syntax.

[7] There is evidence that suggest that the Rhythm Rule is partly—if not wholly—a perceptual phenomenon. Tomlinson et al. (2013), for example, demonstrate that stress shift is still perceived by the listener even when it is not present in the auditory signal. See also Grabe & Warren (1995).

[8] The Rhythm Rule is not always considered one of stress shift (e.g., Gussenhoven 1991).

In (3), *eighteen* exhibits stress shift to the preceding syllable when it is immediately followed by a stress-initial word in the same phrase. In stand-alone form, *eighteen* would usually carry primary stress on the second syllable instead of the first. Kelly and Bock (1988) demonstrate experimentally that pseudo-words in English also exhibit productive stress shift in environments of irregular rhythm.

The third domain of compensation strategies is morphosyntactic adjustment, involving adjustments that utilize constituents in the morphology and syntax—morphemes, words, phrases—in serving the achievement of optimal rhythmicity. Types of repairs in this category include ineffability (blocking), (re-)ordering, and periphrasis, as most commonly noted in the morphological literature (see Chapter 2 for an overview). Dutch plural suppletive allomorphy, as cited in (2) above, is an example of such a morphosyntactic response to rhythmic conditioning. An example of a rhythmically-conditioned syntactic repair comes from binomial pair ordering in English, as noted by McDonald et al. (1993; see also Wright et al. 2005; Benor & Levy 2006; et seq.). In binomial ordering, given the two possible word order candidates (e.g., (4)), previous work has shown that the preference of ordering is one that optimizes rhythmic alternation.

(4)	a.	**péas** and **cár**rots	>	**cár**rots and **péas**
	b.	**áp**ples and **ó**ranges	>	**ó**ranges and **áp**ples[9]

As the binomial pairs in (4) demonstrates, the ordering that optimizes binary alternation between stressed syllables is preferred to the ordering that exhibits a greater irregular lapse pattern between stresses. In the case of (4b), a lapse of two unstressed syllables is preferred to the alternative that features a more severe lapse of three unstressed syllables.

The focus of this thesis is the morphosyntactic "compensation strategies" that satisfy rhythmic well-formedness—in particular, the syntactic adjustments such as periphrasis and word ordering that can optimize for phonological structure. Given that

[9] in dialects where *oranges* is tri-syllabic.

the Principle of Rhythmic Alternation is a constraint assessing a phonological property, the presence of phonetic and phonologically-based strategies in achieving rhythmic well-formedness are naturally expected. And indeed, both phonetic and phonological adjustments for ill-formed rhythmic structure have received a great deal of attention in the existing literature. On the other hand, morphosyntactic adjustments—and especially syntactic adjustments—have been hitherto more rarely discussed. However, that rhythm is a syntagmatic property that applies in contexts where words and larger phrasal constituents combine raises the possibility that morphosyntactic operations might serve rhythmic optimization. This thesis offers evidence that both lexical and syntactic operations optimize for rhythmic alternation.

1.2 Rhythmic optimization and the phonology-syntax interface

Examining the influence of rhythmic well-formedness on the outcome of morphosyntactic processes brings us to the architectural question of the nature of the relationship between phonology and morphosyntax. One extreme view of the linguistic system maintains a feed-forward hypothesis: linguistic derivations proceed through grammatical modules from syntax to morphology to phonology and finally to phonetics and the speech stream. On the other extreme end of the spectrum is the hypothesis that imposes no ordering (and possibly no modularity). Taking seriously the possibility that phonological conditions can impact morphosyntax, the open question is whether there is backwards information flow (“feedback”) or simultaneously-available information between the various parts of grammar.

One common assumption followed by feed-forward models of grammar (e.g., in formal models: Chomsky 1995, et seq.; in psycholinguistic models: Levelt 1989; 2001; Bock & Levelt 1994; Garrett 2000; Ferreira & Slevc 2007) is that of “Phonology-free Syntax” (Zwicky & Pullum 1986a; Vogel & Kenesei 1990; a.o.). The Principle of Phonology-free Syntax disallows forward knowledge of phonological information in the syntactic component of grammar. Because of the one-way flow of linguistic in-

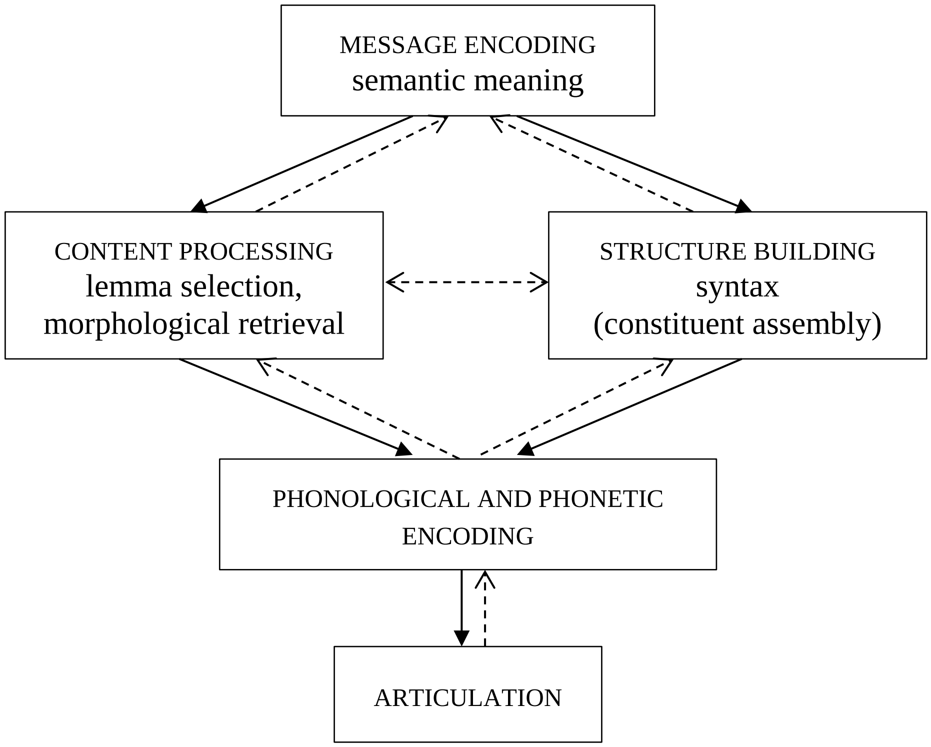

Figure 1. "Consensus model" of language production. Solid arrows represent primary direction of information flow. Dotted arrows represent possible (but debated) direction of information flow. (adapted from Ferreira and Slevc 2007:454)

formation from syntax to phonology, as schematized in Figure 1,[10] phonological information should have no role in syntactic encoding. The oft-trotted-out token example of an illicit rule of phonology in syntax is given by Zwicky & Pullum (1986a:75, their emphasis): "a movement transformation that obligatorily moves… [a] constituent that begins *phonetically* with a bilabial consonant." Phonology-free Syntax insures that such phonetically- (and phonologically-) conditioned syntactic rules do not exist.

Some recent work, however, has argued that phonological information is capable of triggering operations of larger constituents such as words and phrases. The most widely discussed contribution of phonology along these lines is the proposal of a prosodic interface with syntax, where hierarchical prosodic structure (e.g., prosodic

[10] "Consensus model" is Ferreira & Slevc's cover term in summarizing the generally-shared views of the most recent prominent approaches to grammatical encoding in psycholinguistics.

phrases) translates syntactic structure for interpretation by lower-level metrical and segmental phonology (Selkirk 1984; Inkelas 1990; Zec & Inkelas 1990; a.o.). In this way, word order can be prosodically-conditioned: for example, it has been argued that end weight phenomena are conditioned by the Nuclear Stress Rule (Zubizarreta 1998; Anttila et al. 2010; see also Zec & Inkelas 1990; though cf. e.g., Grafmiller & Shih 2011 for a comparison of alternative hypotheses for end weight). Certain cases of this sort where word order is the consequence of prosodic factors that move prosodically-defined constituents to prosodically-defined positions, largely under focus-oriented and discourse-related goals, are termed "PF-movement" (e.g., Zubizarreta 1998; Agbayani & Golston 2010; Agbayani et al. 2011).

Opening the door to prosodic influences on morphosyntactic constituents and operations effectively opens the door to considering the degree to which information from phonology (as well as phonetics) affects morphosyntactic constituents and operations: is phrasal prosody the only phonological information that has a feedback interface with syntactic encoding, or does lower-level metrical and segmental information also affect morphosyntactic operations? A growing amount of evidence has shown that there are not only prosodically-conditioned but also lower-level, phonologically-conditioned effects on word choice, word order, and construction choice in syntax, which necessitate interaction between phonological form and syntactic structure (e.g., Golston 1995; Schlüter 2005) (see more extensive discussion in Chapter 2). Similar effects have also been demonstrated in psycholinguistic studies that feedback exists in the relationship between phonological and morphosyntactic encoding (e.g., Bock 1987; McDonald et al. 1993). Furthermore, parallel questions about the phonology-morphology interface arise in morphology proper. For example, in cases of phonologically-conditioned suppletive allomorphy (e.g., Paster 2006) or phonologically-conditioned ineffability (e.g., Orgun & Sprouse 1999), conditions on the output of phonology influence morphological processes (see also e.g., Kiparsky 1982; 1985).

Regardless of theoretical model, one serious and basic consequence of Phonology-free Syntax has been the position that phonologically-conditioned syntactic phenomena are not empirically attested and the resulting assumption that they are not possible in natural language (cf. Pullum & Zwicky 1988[11]). Due to this assumption, effects of phonological interference with syntax—however small—are not usually not acknowledged or even investigated: only a comparatively small minority of literature has focused on uncovering these phenomena (see Chapter 2 for an overview). Nevertheless, they are crucial information in understanding the architecture of grammar, and the goal of this thesis is to offer more empirical evidence towards considering whether and to what extent the optimization of phonological well-formedness reshapes our fundamental assumptions of how the components of language can interact.

1.3 Thesis overview

This thesis examines the optimization of rhythmic well-formedness in syntagmatic contexts across words and argues that high-level operations in domains beyond phonology and phonetics contribute to phonological well-formedness. While the focus here remains on rhythm as an example of phonological well-formedness conditions, a corollary of the work is the hypothesis that the effects of rhythmic conditions on morphosyntax uncovered here will also extend to other conditions that regulate phonological well-formedness in syntagmatic configuration:[12] that is, morphosyntactic adjustments should be sensitive to not only rhythmic but other phonological constraints. The aim of this thesis is to understand the potential role of morphosyntax in optimizing phonological structure, which should shed light on the relationship between phonological and morphosyntactic encoding at the interface.

---

[11] Pullum & Zwicky 1988 allow for gradient patterns of phonologically-conditioned syntax, dismissing these non-categorical patterns as external to grammar. The approach I take to grammar here, however, is one in which gradient and categorical patterns should arise from the same grammatical principles. See §2 for more discussion.

[12] Indeed, a few such cases will be noted throughout the thesis. See e.g. Obligatory Contour Principle (OCP) effects in lexical choice (Chapter 3) and in construction choice (Chapter 4).

Previously-identified cases of morphosyntactic optimization of phonological constraints are first summarized in Chapter 2. These cases demonstrate that neither metrical nor segmental effects on morphosyntax are entirely unusual. However, the previous evidence of syntactic optimization is largely limited to word reordering or lexical optionality phenomena. In contrast, morphological optimization of phonological well-formedness boasts a broader range of previously-noted repair types, including suppletion, in addition to reordering and blocking.

Turning to the novel contributions of the current work, two case studies of the influence of rhythm in morphosyntactic variation are presented in Chapters 3 and 4. The first—on lexical choice in English name pair formation (e.g., *Susan Smith* versus *Suzanne Smith*)—demonstrates that rhythmic well-formedness, in addition to other phonological preferences (e.g., OCP), is active in the selection of lexical items when two words combine to form a (phrasal) constituent (Chapter 3). The second case study, presented in Chapter 4, examines phonological influences on genitive construction choice in spoken American English (e.g., *the car's wheel* versus *the wheel of the car*). This study finds that phonological rhythm has a role in determining syntactic variation between two alternative constructions for expressing a genitive relationship. Because construction choice cannot merely be linearization of word order, difference in actual syntactic structures must be involved. These two case studies provide empirical evidence that goes beyond previously-discussed, typical word order or lexical optionality effects (see overview in Chapter 2) to show that higher-level morphosyntactic operations can be co-opted to satisfy phonological conditions.

The focus here on rhythmic constraints demonstrates that phonological information at the low levels of the prosodic hierarchy (i.e., rhythmicity across stressed and unstressed syllables)—crucially, *not* hierarchical organization (i.e., prosodic phrasing or cumulative phrasal stress) (cf. Zec & Inkelas 1990:366)—can be involved in the phonological conditioning of morphosyntax. Demonstrating that such low-level phonological constraints (i.e., rhythm; also segmental constraints, though not the focus here) can influence morphosyntactic operations raises the issue of depth of access: how much rhythmic (phonological, phonetic) information is available at the interface?

The latter half of the dissertation turns to this issue by engaging the null hypothesis that all phonetic, phonological, and morphosyntactic information is at once available. Two consequences should arise if there are no restrictions at the syntax-phonology interface: [1] morphosyntactic operations should have access to all phonetic and phonological information, and [2] phonetic, phonological, and morphosyntactic repairs should all compete concurrently to satisfy rhythmic optimality.

Chapters 5 and 6 constitute a narrow examination of the stress and rhythmic properties of grammatical (function) versus lexical (content) word categories to address the issue of available information. Word categories are noted to have different phonological encodings for stress: phonological (underlyingly-encoded) and phonetic (surface-encoded) (e.g., Selkirk 1984; Inkelas & Zec 1993), which arise from differences in either lexical access, lexical storage, or morphosyntactic indexation. Given these differences, if the null hypothesis is correct, then phonetically-encoded information (e.g., surface-encoded stress) should be able to influence morphosyntactic operations. Moreover, phonetic, phonological, and extra-phonological compensation strategies for rhythmic optimization should be available all at once.

A phonetic study of the stress differences in word categories (Chapter 5) presents evidence that these stress differences extend beyond a commonly-assumed binary 'content versus function' word division. It is shown that a more fine-grained distinction, based on phonetic evidence, is necessary to model the complex levels of phonological information that potentially interfaces with the morphosyntax.

Given these results, Chapter 6 presents three case studies of the behavior of different lexical categories on rhythm and syntactic choice. First, it is shown that lexical word categories such as content and function words exhibit different behaviors in rhythmic optimization: phonetic adjustments for irregular rhythmic patterns are more often used for function words than for content words. Such a result suggests that compensation strategies for rhythm vary based on lexical category and depth of phonological encoding of stress. The second case study returns to the genitive alternation discussed in Chapter 4 to probe the issue of whether all phonetic, phonological, and extra-phonological repairs are equally likely at the time of rhythmic optimization. Results of this second case study show that phonetic adjustments for non-optimal rhythmic patterning are not available at the time of morphosyntactic choice. The null hypothesis that assumes no directionality can therefore be rejected, as some amount of directionality between morphosyntax and phonetics must still be present. Finally, the third case study examines whether any phonological information is present in the morphosyntactic candidates that compete to satisfy phonologically-conditioned constraints. Using tests of end weight effects (e.g., the tendency for longer constituents to occur at the ends of phrases), the results of this case study suggest that lexical phonological information about stress is available for morphosyntactic optimization. This is true even when the phonological information is fine-grained, as found in the results reported in Chapter 5.

An ancillary theme that runs through this entire thesis is that rhythmic conditions on morphosyntactic operations must also compete with other factors, including syntactic, semantic, processing-based, and sociolinguistic pressures. The view that emerges is that a variety of conditions—of which phonology is one—contribute to the output of morphosyntax. It is furthermore shown that the effect of rhythm closely interacts with these other factors, such as semantic information (see also e.g. Hanssen et al. 2013). Such complex relationships between phonology and other components of grammar and of usage are likely to be the source of the relative small effect of phonological influences on morphosyntactic operations.[13] That is, in the assessment of morphosyntactic competitors, rhythm has to compete with other active, non-phonological constraints.

The empirical studies presented in the thesis point to a view that allows more phonology-morphosyntax co-dependence than is currently widely accepted. Morpho-

[13] The small influence of phonological factors, as compared to certain higher-order syntactic and semantic factors (e.g., animacy in the genitive construction; see Chapter 4), likely also contributes to their rarity in the previous literature. In fact, such phonological influences on syntactic operations may be common, but heretofore difficult to detect without quantitative observation across sizeable data sources. Some phonological influences, however, have proved to be far stronger and, consequently, prominent in the literature: for instance, adjacent sibilant avoidance in English (Quirk et al. 1985; Zwicky 1987; a.o.) (see also Chapters 3 and 4).

syntactic operations can optimize the well-formedness of phonological structure. Constraints on such interactions are still in place, however, in that morphosyntactic choices are made without competition with phonetically-repaired surface candidates. Chapter 7 concludes the thesis.

The remainder of the current chapter introduces preliminary general background. Section 2 discusses briefly the assumptions behind data, methodology, and theory that are taken in this thesis. Section 3 provides additional background notes on the technical details of rhythm, and §4 includes an overview of the functional motivations of rhythmic well-formedness in language.

2 A note on natural language data, variability, and grammar

The bulk of this thesis reports on variable patterns in morphosyntax, using data from natural (written and spoken) language use. The empirical studies use multivariate, quantitative analyses to investigate the overall and individual contributions of the factors of interest, and also to control for other known variables. The goal in utilizing such data and methodology is to provide quantifiable evidence elucidating the relationship between phonology and morphosyntax in grammatical encoding.

A full discussion of the place of gradience and variability in grammar is beyond the focus of the current work and is not attempted here; however, I will make a few basic remarks.

Historically in theoretical linguistics, the introduction of Chomskyan generativism established a strict division of labor between ‘competence’ and ‘performance’ in language (Chomsky 1957; 1965; 1986; et seq.). In practice, ‘core grammar’—‘competence’—accounted for categorical and absolute grammaticality, while the study of variable and gradient language behavior was relegated to the periphery of ‘language use’ (i.e., ‘performance’) (cf. e.g., Labov 1969; see Wasow 2009 for an overview). Traditional defenses of the Principle of Phonology-free Syntax have also appealed to the strict distinction between categorical 'competence' grammaticality and gradient 'performance'-based variation: in these defenses, apparent cases of phonological influence in syntactic operations such as prosodically-driven NP-shift (see Chapter 2) are excused as language use outside the purview of grammatical responsibility (Zwicky & Pullum 1986b; Miller et al. 1997).

In this work, however, I take as a starting point instead the view that a model of grammar should capture both categorical contrasts and variation observable from natural language use. Such a view has seen an upsurge in recent approaches to phonology,[14] which take variable patterns to arise from the same grammatical systems that generate categorical contrasts (Anttila 1997; 2002; Zuraw 2000; 2010; a.o.). It has been shown extensively in these works that variability—like categorical contrasts—is a patterned phenomenon that arises from principled interaction between established grammatical parameters. Though more controversial, probabilistic approaches based on variable natural language data are also increasing beyond the phonological domain, in morphology (e.g., Hay & Baayen 2005) and syntax (e.g., Manning 2003; Bresnan et al. 2007; Sag & Wasow 2011). Furthermore, probabilistic views have long-standing traditions in sociolinguistic work on language variation and change (e.g., Labov 1972; et seq.), as well as in usage-based approaches (e.g., Bybee 2001; see Bod et al. 2003 for an overview of probabilistic linguistics across subfields). Behind these probabilistic approaches is the tenet that because variation abounds in speech, it represents ecologically-valid data for the study of language that insulates us from potentially biased reliance on non-contextual grammaticality judgments as linguistic data (Gibson & Fedorenko 2010; a.o.). Moreover, recent work has promoted the idea that variation in language actually helps—rather than hinders—language perception and processing (Johnson & Mullenix 1997; Sumner 2011; Sumner & Kataoka 2013; a.o.). Thus, to be able to model the parameters that determine variable patterns in language should be advantageous in understanding linguistic competence overall.

[14] Although this trend coincides with the advent of Optimality Theory and stochastic treatments thereof, it is by no means limited to Optimality-theoretic phonology.

3 Some technical notes on rhythm: A sketch of metrical theory

Formally, rhythmic alternation can be represented using the metrical grid (Liberman 1975; Liberman & Prince 1977; Prince 1983; Selkirk 1984; Hayes 1995; amongst numerous others; and following Lerdahl & Jackendoff 1983 for use of the grid in representing musical rhythm). An example of the metrical grid of the word *supercalifragilisticexpialidocious* is given in (5).

(5)

												*	
*		*		*		*		*		*		*	
*	*	*	*	*	*	*	*	*	*	*	*	*	*
su	per	ca	li	fra	gi	li	stic	ex	pi	a	li	do	cious

In the metrical grid,[15] rhythmic alternation is represented by the regular presence and absence of a gridmark (*) at each level: that is, to be rhythmic, gridmarks on any given level (excepting the lowest) must be separated by at least one gridmark on the level immediately below.

Under the Principle of Rhythmic Alternation, patterns that result in irregularities in rhythmic alternation are dispreferred. Irregularities in rhythmic alternation fall in two categories: lapse and clash. In a lapse situation, stressed syllables are spaced too far apart, as in *míllington's regrét*, shown below.

(6)

*				*
*	*	*	*	*
mil	ling	ton's	re	gret

[15] The grid also represents hierarchical structure, another foundational property of metrical phonology. However, hierarchical structure will not be the focus of the dissertation. For discussion of the relationship between hierarchical properties of rhythm and potential interactions with morphosyntactic operations, see for example Anttila et al. (2010).

In a clash situation on the other hand, an insufficient number of unstressed syllables intervene between (adjacent) stressed syllables, as in the final two syllables of *tènnes-sée áir*:

(7)			*	*
	*		*	*
	*	*	*	*
	ten	ne	ssee	air

Structural constraints on the well-formedness of the metrical grid allow for easy representation of clash and lapse constellations where gridmarks are not separated by exactly one gridmark on the level immediately below.

The grid encodes not only rhythmic regularity but also another property of metrical well-formedness: hierarchical structure. That stress is hierarchical means that there can be *n*-ary levels of differing stress: primary, secondary, tertiary, unstressed, and so on. The grid represents the differing values by levels, and in a well-formed grid, rhythmic alternation should hold at all levels (Hayes 1995:28). The example in (8) demonstrates rhythmic clash at the word level because the gridmarks are not separated by at least one gridmark in the foot level immediately below, despite being separated by a gridmark two levels below in the syllable layer:

(8)			*		*	word
	*		*		*	foot
	*	*	*	*	*	syllable
	mi	ssi	ssi	ppi	mud	

The primary focus of this thesis will be the property of rhythmic regularity in metrical well-formedness. Concentrating on rhythmic regularity demonstrates that phonological properties other than the hierarchical organization of prosodic structure can affect morphosyntactic operations (cf. Zec & Inkelas 1990:366). For the purposes of the discussion here, the example in (8) is rhythmic—that is, has absolute alternation between strong and weak elements. For simplicity, the differences between levels of stress (e.g., primary versus secondary) will not be explored, but this does not mean that they have no potential influence on influencing morphosyntactic operations (e.g., Anttila et al. 2010 suggest that primary and secondary stress may contribute differences in weight). The role of hierarchical structure in assessing metrical well-formedness, if any, is saved for future work.

4 Functional advantages of rhythmic regularity

Rhythmic regularity is not a vacuous formal construct: its existence serves language production, planning, and perception. For example, Tilsen (2011) showed experimentally that the recall accuracy and speed of rhythmically regular sequences were greater than rhythmically irregular ones, suggesting that regularity eases speech planning and production. Because the presence of rhythm creates expectations of regularly distributed strong and weak elements, Large and Palmer (2002) proposed that more auditory attention is paid to events occurring in strong positions due to heightened expectations, giving strong beats a perceptual advantage. Their theory has been corroborated with evidence in the realms of music perception (Repp 2010) and language perception (Tomlinson et al. 2013).

Beyond the individual speaker, rhythmic regularity and expectation can also play into conversational coordination and the social dynamics of language. Uhmann (1996) provides evidence from German spontaneous conversation that speakers manipulate rhythmic expectations for emphasis, exaggeration, and other extra-linguistic meanings. In the conversation analysis literature, rhythmic synchronization between speakers has been proposed to account for the coordination of turn-taking (Couper-Kuhlen 1993; Auer et al. 1999; a.o.).

Additionally, work with aphasic subjects has shown that phonological constraints such as rhythm are strong enough to permeate into tautomorphemic environments. In Cohen-Goldberg et al.'s (2013) study, their aphasic individual, who had trouble producing phonologically-complex multi-morphemic constructions, upheld metrical constraints across morpheme boundaries. The individual routinely avoided stress clash by inserting epenthetic syllables between morphemes: *briskness* → [brɪsk**ɪd**nɛs].

Given the evidence from previous research, rhythmic regularity clearly has a strong functional basis in language and cognitive processes.

Chapter 2. Phonologically-conditioned morphosyntactic phenomena

1 Introduction

The focus of this thesis is the potential for morphosyntactic processes such as lexical, word order, and construction choices to optimize for phonological well-formedness. The impact of phonological conditions on *syntactic* phenomena may be regarded as a somewhat radical departure from standard views of how phonology-syntax interact (see discussion in Chapter 1) (see also discussions in e.g. Zec & Inkelas 1990; Schlüter 2005; Jaeger 2006; cf., Zwicky & Pullum 1986b; Miller et al. 1997). The impact of phonological conditions on *morphological* phenomena, however, constitutes a long-standing issue in the study of morphology. This chapter outlines the empirical parallels between phonologically-conditioned morphology and previously-noted instances of phonologically-conditioned syntactic phenomena. Viewing the influence of phonology in the domain of syntax as a natural counterpart to the more familiar influence of phonology in the morphological domain results in typological implications for considering the range of phonological-conditioning across morphosyntax at large. In particular, the aim of this chapter is to demonstrate that phonological interference with syntactic phenomena is not as unexpected or as unprincipled as it may be widely perceived to be.

This chapter reviews known cases of phonological influences on morphosyntactic operations. Section 2 presents a brief survey and working typology of phonologically-conditioned morphology. The remaining bulk of the chapter focuses on syntactic effects across words and demonstrates the parallels between phonologically-conditioned morphology and phonologically-conditioned syntax (§3ff). The syntactic phenomena range from those triggered by prosodic (§3.1) and metrical (§3.2) conditions to those triggered by segmental and syllable structure (§§3.3–3.4) conditions.

This list of phonologically-conditioned morphosyntactic phenomena is by no means exhaustive.[16] Section 4 concludes.

2 Phonologically-conditioned morphology: a working typology

The influence of phonology on morphological processes is by no means an uncontroversial issue (see e.g., Nevins 2011; Inkelas 2014: Chapter 9 for overviews). For the current purposes, however, I focus on the empirical phenomena and present a generalized, working typology of the various phonologically-conditioned morphological phenomena that have been reported. Because the goal here is to point to parallels between empirical behavior in phonologically-conditioned morphological and syntactic effects, I remain agnostic with regards to the theoretical approaches and ramifications of these phenomena (cf. Carstairs 1990; McCarthy & Prince 1993a; 1993b; Mester 1994; Paster 2006; 2009b; Wolf 2008; Embick 2010; a.o.).

There are two components to consider in phonologically-conditioned morphosyntactic effects: the phonological conditions (§2.1) and the morphosyntactic responses (i.e., "repairs") to phonological conditioning (§2.2). These are discussed in turn in the following sections.

2.1 Phonological conditions

The phonological conditions that can influence morphology range from prosodic and metrical well-formedness (e.g., *CLASH) to constraints regulating segmental patterns (OCP). Representative examples are given below. The issue of whether these conditions must be phonologically optimizing (as opposed to phonologically arbitrary) will

[16] For surveys of phonologically-conditioned morphology, see Carstairs-McCarthy (1998); Paster (2006); Nevins (2011); Inkelas (2014); a.o. For a survey of rhythmically-conditioned syntax, see Schlüter (2005).

not be taken up here; but, I will note that it is a crucial issue that informs the debates surrounding phonology-morphology interaction (see Booij 1998; Paster 2006; 2009b; Wolf 2008; Embick 2010).

2.1.1 Prosodic conditioning

Clitic placement in Serbo-Croatian is one example of prosodically-conditioned morphology. In Serbo-Croatian, second position clitic placement is variable, either following the first syntactic phrase [*veoma važan*]$_{AP}$ (9a) or following the first prosodic word [*veoma*]$_{Pw}$ (9b) (examples from Zec & Filipović-Đurđević, to appear).

(9) a.

Veoma	važan	je	taj	zadatak	
very	important		is-CL	this	task

‘This task is very important.’

b.

Veoma	je	važan	taj	zadatak	
very	is-CL	important		this	task

‘This task is very important.’

Zec & Filipović-Đurđević (to appear; following Inkelas 1990; Zec & Inkelas 1990; Halpern 1992; a.o.) argue that the latter context must be in part prosodically conditioned,[17] because the prosodic word does not always coincide with boundaries of syntactic constituents:

(10) a.

Na	crnom	je	stolu	ostala	samo	jedna	knjiga
on	black	is-CL	table	remain	only	one	book

‘Only one book remained on the black table.’

[17] Zec and Filipović-Đurđević also argue that the clitic placement cannot be solely prosodic, as information about the larger syntactic frame (i.e., predicate- versus argument-initial structure) factors into the acceptability of clitic placement after first prosodic words in Serbo-Croatian.

b.	Na	crnom	stolu	je	ostala	samo	jedna	knjiga
	on	black	table	is-CL	remain	only	one	book

‘Only one book remained on the black table.’

Example (10a) demonstrates a case where the first prosodic word—the host of the clitic *je*—is made up of a preposition and adjective [*na crnom*]$_{Pw}$, which do not together form any syntactic constituent (cf. (10b), in which [*na crnom stolu*]$_{NP}$ is a noun phrase). Thus, the data suggests that prosody may be in part responsible for defining the constituency needed for determining clitic placement.

Other known prosodic conditions include foot structure optimization and prosodic minimality restrictions (see summaries in Paster 2006:140ff; Nevins 2011:12–14), with a large literature in prosodic morphology that derives morphological patterns (e.g., reduplication, clipping, infixation) from prosodic considerations (McCarthy & Prince 1993b; et seq.)

2.1.2 Metrical conditioning

The avoidance of stress clash (*CLASH) and lapse (*LAPSE) has been noted to elicit a number of morphological responses. As demonstrated in Chapter 1, Dutch plural suppletive allomorphy is conditioned by the avoidance of stress lapse. In English, the ineffability of certain suffix-root combinations has been attributed to the avoidance of stress clash (Raffelsiefen 1996; 1999; Plag 1999; Smith 2013; a.o.). For example, the suffix *–ize* may acceptably attach to words with penultimate stress *schéma* → *schématìze*. Stress-final words, however, are blocked from *–ize* suffixation due to the potential result of stress clash: *secúre* + *–ize* → **secúrìze*.

2.1.3 Syllable structure conditioning

In addition to metrical considerations, other suprasegmental conditions have been demonstrated to affect morphology: syllable structure optimization, phonological and morphological alignment (e.g., Nevins 2011:6–10 for an overview), and tone (e.g., Mortensen 2006:235ff; Paster 2006:127–131). One of the most common conditions of this type is syllable structure optimization. In Korean, for example, suppletive allomorphy between nominative case suffix allomorphs –*i* and –*ka* depend on satisfying syllable structure well-formedness (example from Paster 2006:67; citing Odden 1993):

(11)	a.	param-**i**	'wind-NOM'
	b.	pori-**ka**	'barley-NOM'

As shown in (11), the allomorph –*i* follows consonant-final roots and the allomorph –*ka* follows vowel-final roots. This pattern achieves optimal CV syllable structure by avoiding coda consonants (*CODA; e.g., **para**m**-ka*) and hiatus between two vowels (ONSET; e.g., **pori**-i**).[18]

2.1.4 Segmental conditioning

Segmental conditions that elicit morphological responses are generally ones that regulate phonotactic preferences between two segments. A well-cited example of this type is the co-occurrence restriction militating against adjacent sibilant sequences in English: ineffability will often result when illicit sibilant-sibilant sequences are encountered, as in **two McDonald's-s* (Menn & MacWhinney 1984; Zwicky 1987; beyond sibilants, see also Yip 1998). Sibilant co-occurrence restriction also triggers suppletive

[18] See Haitian Creole articles for a pattern that runs counter to Korean syllable optimization (Klein 2003). In Haitian Creole, consonant-initial articles follow consonant-final syllables, and vowel-initial articles follow vowel-final syllables. Sometimes these vowel-vowel sequences are repaired by glide insertion; at other times, vowel hiatus—particularly of low vowels (e.g., [a] – [a])—is tolerated.

allomorphy in Hungarian, for example. Where the second person singular is usually marked by an [-s] allomorph (12a), an [-ol] allomorph appears following sibilant-final verbs instead (12b) (data from Paster 2006:41–42; citing Abondolo 1988; Rounds 2001). Note that [s] = *sz* in Hungarian orthography.

(12)	a.	vág-sz	'you cut'
		vár-sz	'you wait'
		nyom-sz	'you press'
		rak-sz	'you place'
	b.	vonz-ol	'you attract'
		edz-el	'you train'
		hajhász-ol	'you seek'
		főz-öl	'you cook'

Long-distance phonotactics are also known to result in morphological gaps. In Tagalog, a co-occurrence restriction on sonorant labial consonant sequences (e.g., *[m…m], *[w…m]) that may be separated by a vowel results in the blocking of *-um-* infixation (data from Zuraw & Lu 2009:199; see also Schachter & Otanes 1972; Orgun & Sprouse 1999):

(13)		stem	actor focus, infinitive/past	
	a.	pili	p-um-ili	'choose'
		takot	t-um-akot	'frighten'
		kanta	k-um-anta	'sing'
		bukas	b-um-ukas	'open'
		damaj	d-um-amaj	'sympathize'
		gawaʔ	g-um-awaʔ	'make'
		sulat	s-um-ulat	'write'
		nipis	n-um-ipis	'become thin'
		ŋitiʔ	ŋ-um-itiʔ	'smile'
		lipat	l-um-ipat	'move'
		ʔawit	ʔ-um-awit	'sing'
		hiŋiʔ	h-um-iŋiʔ	'ask for'
		jakap	j-um-akap	'embrace'
	b.	meɾi	***m**-u**m**-eɾi	'marry'
		wejl	***w**-u**m**-ejl	'wail'

As shown in (13b), words that would result in a sonorant labial + sonorant labial sequence block -*um*- infixation[19] while other words tolerate infixation (13a). Orgun & Sprouse (1999) demonstrate that this ineffability is productive, and applies to potential new loans, as in (13b). The same co-occurrence restriction (between labial + labial sequences) in other Austronesian languages triggers phonological repairs such as fusion (e.g., in Mayrinax Atayal) and dissimilation (e.g., Limos Kalinga) (Zuraw & Lu 2009; see Coetzee & Pater 2008 for OCP effects in Muna, another Austronesian language).

2.2 Morphological responses to phonological conditioning

Morphological responses to phonological conditioning—"repairs", if phonologically-optimizing—typically take one of the following forms:

1. Blocking (ineffability)
2. Periphrasis
3. Ordering
4. Suppletion (or, more generally, replacement of one form by another)
5. Deletion (haplology/omission)

Examples of each type of response are provided below.

2.2.1 Blocking

Blocking is not so much a "repair" as a potential consequence of phonological conditioning. If a phonological requirement blocks a morphological configuration from surfacing, then there are one of two possible outcomes: [1] either a repair to satisfy the

[19] Zuraw & Lu note two [w]-initial exceptions in their Tagalog dictionary data (from English 1986).

phonological requirement takes place in order to allow a licit surface form, or [2] no repair is utilized and the underlying communicative intent remains unexpressed.[20] The latter instance would be the extreme outcome of blocking—true ineffability.

Evidence for ineffability comes largely from the presence of (paradigm) gaps. The blocking of Tagalog -*um*- infixation (see §2.1.4), for example, causes gaps in the paradigm in cases of labial-labial co-occurrence restriction violations (e.g., ****m***-*u****m***-*eɾi*; cf. example (13)). Gaps created by phonologically-conditioned blocking are sometimes repaired via another strategy (e.g., periphrasis, suppletion), though these repairs are not always systematic (Carstairs-McCarthy 1998). For instance, the gap resulting from sibilant OCP blocking of *–sz* suffixation in the second person singular paradigm of Hungarian (see (12)) is systematically filled by the use of suppletive allomorph *–ol*, which originated historically from a separate inflection class (Carstairs-McCarthy 1998:147). On the other hand, no systematic (either morphological or periphrastic) alternative for gaps resulting from labial OCP in Tagalog is reported. Carstairs-McCarthy notes that this disjunction between systematic reparations for blocking-induced gaps coincides with a division between inflectional versus derivational morphology: observationally, gaps in inflectional paradigms are more often systematically filled.

2.2.2 Periphrasis

Periphrasis, broadly interpreted, is the use of one construction—usually analytic (i.e., made up of multiple words)—to expone the same semantic material as another construction—usually synthetic (i.e., contained in a single word). For example, the English comparative suffix *–er* alternates with a periphrastic competitor that uses *more* + adjective, producing variation between synthetic and analytic forms: *yellow-er* versus *more yellow*, respectively. This variation is prosodically-conditioned, in addition to

[20] assuming, for simplified, expository purposes, that communicative intent is the underlying form; such an assumption has some traction in phonological analysis: see e.g., Zuraw (2000); Kiparsky (2005).

being affected by lexical frequency (Poser 1992; Kiparsky 2005; Embick 2007; Embick & Marantz 2008; Adams 2014; to appear; a.o.). The general observation is that shorter words are more likely to allow *–er* suffixation, as in the disyllabic example (14a). Long words are less likely to tolerate suffixation, as in the quadrisyllabic example in (14b). In instances where the *–er* suffix is likely blocked due to prosodic (or frequency) conditions, the analytic alternative is instead preferred.

(14)

	adjective	synthetic form		analytic form
a.	happy	happier	*more likely than (>)*	more happy
b.	salubrious	?salubriouser	*less likely than (<)*	more salubrious

Poser (1992) hypothesizes that such periphrastic alternations with morphological categories are limited to "small categories" in the syntax. Under Poser's proposal, only categories that dominate zero-level projections are eligible to block or be blocked by synthetic forms.

It should be noted here that a more general form of periphrasis—paraphrasing—is always a viable repair option, especially if the desire to express a communicative intent is sufficiently strong.[21] In fact, the line between periphrasis and paraphrasing is a thin one. The two concepts are separated perhaps only by the understanding that periphrasis involves regularized and systematic alternatives of forms or rules. Paraphrasing, on the other hand, has no such restriction: its only constraints are satisfying expressiveness and economy. The consideration of paraphrasing as a general type of periphrasis repair is important in the discussion of parallels between phonologically-conditioned morphology and syntax because it is possible that paraphrases can be legitimate competitors with synthetic forms in comparative grammaticality at a sentential level.

[21] assuming, again, that communicative intent is the input (see fn 20). The "desire to express communicative intent" can be formalized m a number of ways: for example, in Optimality-theoretic terms, either a version of MPARSE (Prince & Smolensky 1993) or a version of faithfulness—e.g., EXPRESSIVENESS ("Express meaning.") (Kiparsky 2005:114).

2.2.3 Ordering

The ordering of morphemes can be phonologically-conditioned. Variable clitic placement (as discussed in §2.1) and mobile affixes (e.g., Kim 2010) are examples of phonologically-conditioned morpheme ordering. Compound formation is another. Mortensen (2006) presents a number of coordinate compound ("co-compound") ordering phenomena that depend on phonological constraints regulating long-distance phonotactic (e.g., vowel height patterns) and tonal preferences. In Jingpho, for example, a preference for high vowels followed by low vowels determines the order of elements in co-compounds (example from Mortensen 2006:222–223):

(15)	a.	lùʔ	+	ʃá
		drink		eat
		'food'		
	b.	*ʃá	+	lùʔ
		eat		drink

As (15) illustrates, ordering between two elements will select the co-compound in which the high vowel [u] occurs before the lower vowel [a]. The order that results in the opposite vowel height pattern is ungrammatical (15b). While in Jingpho, vowel height preferences produce a categorical order distinction, it is important to note that the same high-to-low vowel quality constraint can also result in variable lexical ordering patterns, as in Lahu co-compounds (Mortensen 2006:215–221). This demonstrates that the same phonological condition can underlie both categorical and variable patterns, even for phonologically-conditioned morphological phenomena.

Another example of phonologically-conditioned morpheme re-ordering comes from Slavey (Rice 2011:183–184), in which affixes with less phonological material (e.g., possessive suffix -*ɛ́*)—that is, "light" suffixes—surface closer to the stem than affixes with more phonological material (e.g., diminutive -*zha*)—i.e., "heavy" suffixes:

(16)	a.	-’ah- έ’	‘snowshoes-possessed’
	b.	-’ah- **έ**-zha	‘small snowshoes, women’s snowshoes’
	c.	*-’ah-zha- **έ**’	

Rice notes that the ordering of the lighter, possessive suffix inside the heavier diminuitive in (16b) flouts the ordering predicted by scope (16c), which would otherwise place the possessive suffix outside of the diminuitive.

2.2.4 Suppletion

Suppletive allomorphy is the alternation between two or more distinct and phonologically unrelated surface forms (in complementary distribution) that expone the same semantic material. Suppletive allomorphs are typically sufficiently different in terms of phonological form so that they cannot be attributed as derived from a shared underlying form. Suppletive allomorphy of this type can be phonologically-conditioned (commonly referred to as “PCSA” for Phonologically-Conditioned Suppletive Allomorphy, following Paster 2006). PCSA examples include the Hungarian second person singular allomorphs –*sz* and –*ol* that are conditioned by OCP, as presented in (12), and Dutch plural allomorphs –*s* and –*en* that are conditioned by *LAPSE, as presented in Chapter 1.

2.2.5 Deletion

Phonological conditions can often result in the deletion of phonological segments that make up the entirety of a short affix (cf. haplology of an entire morpheme; see e.g., Menn & MacWhinney 1984 for an overview of phenomena). In English, for example, the avoidance of adjacent sibilants (as mentioned in §2.1.4) can cause the deletion of one of the sibilants, particularly when the plural and possessive suffixes are both attached: *student+ -s + ’s → students’* [st^{j}udn̩ts], *st^{j}udn̩ts-əz.

3 Phonologically-conditioned syntax: an overview

In this section, I compare the phenomena of phonologically-conditioned syntax with phonologically-conditioned morphology, as overviewed in §2 above. The question at hand is whether and to what extent phonology-syntax effects are of the same type as phonology-morphology interactions. Phonologically-conditioned syntactic phenomena that have been noted in the previous literature are summarized in this section. I suggest here that similar types of repairs are involved: in particular, periphrasis, re-ordering, and deletion of syntactic material are common responses to phonological conditioning. Furthermore, the phenomena are affected by similar prosodic (§3.1), metrical (§3.2), syllabic (§3.3), and segmental (§3.4) constraints that have been shown to condition morphological effects. The crucial difference between phonologically-conditioned morphology and phonologically-conditioned syntax instead seems to lie in the strength of the interactions across morpheme and word boundaries: syntactic phenomena exhibit weaker and often more gradient effects. The discussion in §4 will return to this observed difference.

3.1 Prosodically-conditioned syntactic phenomena

The most widely acknowledged phonological interface with syntax is the high-level prosodic phrase domain, where morphosyntactic processes are triggered by considerations of prosodic structure.

Zec and Inkelas (1990) were amongst the first to posit that heavy constituent licensing is a consequence of prosody. “Heavy” constituents have a strong tendency to appear at phrase-boundaries, being pulled to one or the other periphery. In English, this is known more popularly as the Heavy-Last Principle (Behaghel 1909; Quirk et al. 1985; Wasow 2002; a.o.). Zec and Inkelas hypothesize that dislocated complex noun phrases are licensed only when they are sufficiently heavy, defined prosodically: "heavy" NPs consist of at least two prosodic phrases (ϕ-phrase):

(17) a. Mark showed [some letters] $_{\phi}$ to John.
b. ???Mark showed to John [some letters]$_{\phi}$.
c. ✓Mark showed to John [some letters]$_{\phi}$ [from Paris]$_{\phi}$.

The unacceptability of (17b; cf. 17a) stems from the fact that *some letters* consists of only one prosodic phrase. In contrast, the NP shift in (17c) (cf. *Mark showed some letters from Paris to John*) is tolerated because *some letters from Paris* consists of two prosodic phrases, making the constituent sufficiently heavier as a trigger of shift. In other words, the lightness (versus heaviness) of the prosodic phrase is capable of blocking NP shift.

A related, prosodically-based analysis of heavy NP shift has suggested that the alignment of lexical and phrasal stress under the rightwards Nuclear Stress Rule (Chomsky & Halle 1968) is responsible for the heavy-last condition in English (Zubizarreta 1998; Anttila et al. 2010). In this view, phrase-final nuclear stress attracts stressed elements; consequently, a constituent that contains more stresses will be more likely to shift and receive nuclear stress than a constituent that contains fewer opportunities for nuclear stress docking. Anttila et al. (2010:954) formalize this approach as the STRESS TO STRESS PRINCIPLE, as defined in (18).

(18) STRESS TO STRESS PRINCIPLE: Each lexical stress occurs within the prosodic phrase that receives sentence stress.

By the principle in (18), lexical stresses are attracted to rightmost sentential stress, as dictated by the Nuclear Stress Rule. Returning to the example in (17), *some létters from Páris* in (17b) has a greater number of lexical stresses (n = 2) than *some letters* or *Jóhn* (n = 1). Thus, *some letters from Paris* will be more preferred in the position that receives sentential nuclear stress, thereby licensing NP shift. "Heaviness" here is defined prosodically in terms of lexical stress weighting.[22] As such, a prosodic property conditions syntactic constituent order.

3.2 Metrically-conditioned syntactic phenomena

An extreme instance of rhythmic well-formedness properties that affect syntax is found in verse practices in English. Youmans (1983) noticed that (ungrammatical) syntactic inversions occur in Shakespeare's verse when the inverted order obeys metrical constraints leading to greater rhythmic regularity (see also Fitzgerald 2007). The syntactic inversion of *blunter* and *be* in line (19a) is allowed because the syntactically-grammatical line (19b) would be rhythmically problematic, with the main stress of a polysyllabic content word falling in a metrically weak position,[23] as represented by "w" in the scansion below each line in (19).

(19) a. Thy édge should **blún**ter be than áppetite (*Sonnets* 56.2)
w s w **s** w s w s w s

b. *Thy édge should be **blún**ter than áppetite (construct)
w s w s **w** s w s w s

That poets utilize syntactic re-ordering to achieve rhythmic regularity is hardly surprising in a linguistic art form such as verse; however, insofar as poetic forms are a derivative reflection of the properties of natural language (Kelly & Rubin 1988; Palmer & Kelly 1992; Hanson & Kiparsky 1996; a.o.), the use of syntax to achieve optimal rhythmic patterning in Shakespeare demonstrates the possibility of metrical influences on syntax.

---

[22] The definition of 'weight' as a prosodic or phonological property is not an uncontroversial one. Numerous definitions of weight that depend on a variety of measures, from information content to phonological content, have been proposed. See e.g., Wasow (2002); Szmrecsányi (2004); Grafmiller & Shih (2011) for overviews and comparisons. This issue is also taken up briefly in Chapters 4 and 6.

[23] see Kiparsky (1977) for metrical mapping conditions.

3.2.1 Ordering

Beyond verse, the metrical properties of clash and lapse avoidance have been noted to trigger ordering alternations. McDonald et al. (1993:215) found that rhythmic alternation influences English binomial pair order. In coordinating nouns, experimental subjects preferred to order noun constituents to maximize alternation between stressed and unstressed syllables, when semantic factors like animacy were held constant. For example, subjects were more likely to order the constituents *sin* and *surprise* as the order in (20a) over the alternative ordering in (20b):

(20) a. sur**príse and sín** >
b. sín and surpríse

The ordering in (20a) promotes binary rhythmic alternation between the stressed syllables of the two nouns while the ordering in (20b) incurs a rhythmic lapse of two unstressed syllables between stressed ones.

Corpus studies of binomial ordering have also noted significant lapse-avoidance effects in conditioning variation between pair order (e.g., *compléte and ùnabrídged* > *ùnabrídged and compléte*, as reported by Benor & Levy (2006:243, 253); see also discussion in Wright et al. (2005)). Furthermore, Mollin (2012:97ff) found that frozen binomial pairs—that is, pairs that demonstrate less reversibility in a large corpus (e.g., British National Corpus)—were more prone to following metrical constraints such as lapse avoidance. This suggests that rhythmic well-formedness promotes preference for a given construction, leading to increased frequency and the eventual development of frozen form.

3.2.2 Paraphrasing

Schlüter (2005) also argues that rhythmic considerations impact morphosyntactic changes over time. For example, the avoidance of *a*-adjectives in prenominal positions is hypothesized to be affected by clash avoidance.

(21) a. the **pérson who was awáre**
b. > ?the a**wáre pér**son
c. > ??the a**sléep pér**son

In the example above, (21a) is preferable to (21b, c) because it avoids the stress clash that is incurred when a stress-final *a*-adjective precedes a stress-initial noun. Under Schlüter's view, because most nouns in English are stress-initial, the repeated avoidance of prenominal *a*-adjectives due to stress clash affected the diachronic shift away from adjective-noun order in these cases. The alternative is paraphrasing of the adjective-noun, which can be viewed as a generalized, non-systematic form of periphrasis.

3.2.3 Deletion

Apart from ordering and re-wording preferences, clash and lapse avoidance have been posited to condition the realization of words (usually function words) in speech. Wasow et al. (2012) found that the use of optional *to* in the *do be* construction is in part conditioned by rhythmic factors. Under the assumption that *be* must carry stress because it cannot be reduced (e.g., *All I want to* ****do's***), the optional realization of *to* is shown via a corpus study (i.e., COCA) to be dependent on avoiding stress lapse:

(22) a. All I want to do **ís to gó** to work.
b. All I want to do **ís repórt** my work.
c. > All I want to do **ís to repórt** my work.

In the example above, *to* occurs more often when its appearance promotes rhythmic regularity in relation to the surrounding syllables. If, as in (22c), *to* would create lapse, it is found to be quantitatively dispreferred.

Similarly, Lee & Gibbons (2007) report that the use of optional complementiser *that* is influenced by rhythmic structure. In the examples below (from Lee & Gibbons 2007:450), the realization of *that* depends on its surrounding stress environment: [24]

(23) a. Henry **knéw that Lú**cy washed the dishes.
b. Henry **knéw Louíse** washed the dishes.
c. > Henry **knéw that Louíse** washed the dishes.

If the immediately preceding and following syllables are stressed, then *that* occurs to avoid clash (23a). Otherwise, if there is already an adjacent stressless syllable, *that* is less likely to be produced (23b) because it would incur a greater lapse (23c).

3.3 Syllable structure conditions on syntactic phenomena

Syllable structure and weight are common conditions on syntactic phenomena. In corpus studies of binomial ordering, Wright et al. (2005) and Benor & Levy (2006) find that ordering preferences are affected by both syllable structure well-formedness conditions and syllable count. For example, the ordering of *John and Yoko* is preferred to *Yoko and John* because the latter option violates a purportedly-universal constraint for syllable to have onsets (i.e., ONSET; Prince & Smolensky 1993). *Yoko and John* create a hiatus between *Yoko* and the conjunction *and* whereas the ordering *John and* provides an onset [n] for the following syllable.[25]

---

[24] operating under the assumption that *that* is always unstressed.

[25] Note, however, that allowing [n] to syllabify as the onset of *and* violates alignment of morpheme boundary to syllable boundary (McCarthy & Prince 1993a).

Furthermore, the dispreferred ordering of *Yoko and John* flouts the heavy-last preference in English for long constituents to occur in the rightmost position. Because *Yoko* is two syllables long and *John* consists of only one syllable, *Yoko* is preferred in second position. It is worth noting, however that in a study of reversibility in English binomial pairs, Mollin (2012) reports a dichotomy between metrical and segmental constraints on binomial ordering. Metrical constraints such as rhythm robustly predicted how frozen a binomial pair could be, suggesting a correlation between metrical well-formedness and how likely a binomial pair would propagate in a certain order. Segmental constraints, on the other hand, did not have such an effect.

3.4 Segmental conditions on syntactic phenomena

Conditions on segmental material have also been noted to interact with morphosyntactic operations, even though an empirical consequence of Phonology-free Syntax has been that this type of interaction does not exist.

3.4.1 Ordering

The most commonly-discussed phenomena of segmental phonology and syntax interaction are haplology effects (e.g., Yip 1998 for an overview). For example, Golston (1995:353–354) shows that center-embedded noun phrases in Ancient Greek are blocked when the process would result in a sequence of surface-identical words:

(24) a. [t-ées [t-óon himatí-oon] ergasí-as]
the-G:F the-G:N:P clothing-G:N:P production-G:F
'of the production of clothing'

b. *[[t-óon [t-óon eikeín-oon] oikeí-oon] tin-ás]
the-G:F:P the-G:M:P those-G:M:P slave-G:F:P some-A:F:P
'some of the slaves of those [people]'

c. [[[t-óon oikeí-oon] tin-às] [t-óon ekeín-oon]]
the-G:F:P slave-G:F:P some-A:F:P the-G:M:P those-G:M:P
'some of the slaves of those [people]'

As shown above, center-embedding of a noun phrase is acceptable when the embedding does not produce a homophonous sequence of articles (24a). When embedding would result in a homophonous sequence (24b), it is unattested, and post-posing (24c) is used instead. Such avoidance of a repeated set of morphemes using a syntactic alternative demonstrates that syntactic decisions need to be able to access segmental information from phonology in order to identify and avoid string similarities.

Another haplology effect that is common to English is the avoidance of stacked sibilants (Quirk et al. 1985; Zwicky 1987; et seq.), usually incurred by the plural morpheme or the possessive clitic -*'s* (see also §2.2.5). Quantitative corpus studies of the English genitive alternation have shown that sequences of sibilant and the clitic -*'s* are dispreferred and robustly influence the choice of the alternative genitive construction using *of*: for example, *the chur**ch's** bell* is dispreferred to *the bell of the church*, all else being equal (Hinrichs & Szmrecsányi 2007; Szmrecsányi & Hinrichs 2008; Shih et al. to appear) (see also Chapter 4).

3.4.2 Deletion

Additional haplology effects are also shown to be at work in *that* optionality. Jaeger (2006:9; see also Walter & Jaeger 2005) demonstrates in a corpus study that *that* is less likely to occur when its occurrence results in a repetition of the word:

(25) I heard **that that**'s the second happiest day of your life.

Jaeger does warn that this type of whole lexeme avoidance could be attributed to processing interference rather than phonological avoidance, however; regardless, phonological form is necessary for such an identity effect.

3.4.3 Periphrasis

Periphrasis (more generally, paraphrasing) and re-ordering of syntactic constituents are similar repairs. Periphrasis is broken out from the discussion on ordering effects (§3.4.1) here to highlight phenomena in which the alternatives are two different syntactic surface structures rather than merely a re-ordering of words with no apparent surface differences (e.g., additional function words or phrasal structures). The actual theoretical differences between word order and surface structure are set aside here, but it should be noted that depending on the theoretical approach to syntax, the formal distinction between ordering and periphrasis may only be superficial.

Bock (1987) provides evidence from psycholinguistic priming that phonological information can influence the choice of active or passive syntactic constructions. Given an instance in which *barn* is phonologically primed via prior exposure to a similar onset or vowel, then the active sentence (26a) will more likely be produced, as opposed to the passive sentence (26b).

(26) a. A tornado is hitting a barn.
b. A barn is being hit by a tornado.

Bock attributes this effect to accessibility: more accessible items are produced first. Phonological priming, however, decreases the accessibility of a phonologically-similar item, due to interference with the form of the original word in lexical access. Consequently, syntactic differences demonstrate sensitivity to such phonological difficulty by placing phonologically less accessible words later in the utterance.

4 Conclusion

The goal of this chapter was to present evidence that, to a certain extent, phonologically-conditioned syntactic phenomena can be viewed as macrocosmic extensions of phonologically-conditioned word-internal (or prosodic word-internal) morphological phenomena. A general typology of the empirical phenomena that have been previously identified as phonologically-conditioned morphology is summarized below:

(27)

Phonological conditions	*Repairs/responses*
Prosodic conditions	Blocking (ineffability)
Metrical conditions	Periphrasis (and paraphrasing)
Suprasegmental conditions	Ordering
Segmental conditions	Suppletion (replacement)
	Deletion (haplology/omission)

Morphological repairs and responses of blocking, periphrasis, ordering, suppletion, and deletion have been shown to be triggered by a wide range of phonological conditions, ranging from prosodic and metrical to segmental. One common thread to the phonological conditions that have been found is that they are all ones that mandate preferences over syntagmatic configurations that can be triggered when morphological combinatorics and processes occur (e.g., OCP, phonotactic restrictions).

These same phonological conditions that affect the morphological domain have also been shown to trigger syntactic repairs and responses, as reviewed in §3. In the previous morphological literature, such cross-word effects are commonly decried (e.g., Carstairs 1990:20), but they in fact do occur. Phonological conditions can affect syntactic constituents similarly as they affect material in the morphological domain. Commonly, we see periphrasis (or paraphrasing), re-ordering, and deletion of syntactic material as responses to prosodic, metrical, suprasegmental, segmental conditions.

Some differences do obtain in the empirical survey. For example. phonologically-conditioned affix re-ordering phenomena are rare at best (e.g., Paster 2009a), and most of the instances that I am aware of do not involve metrical conditioning. On the other hand, phonologically-conditioned syntactic re-ordering (of words or larger syntactic constituents) is one of the most common types of phonologically-conditioned syntactic effects noted in the existing literature, and metrical constraints are a common trigger of such re-ordering. It is left for future survey whether this difference is an accident of our empirical knowledge or whether it underlies a deeper, systemic distinction between phonological conditioning of syntax versus morphology.

There are two responses that we see in the morphological literature that are rarer in syntactic phenomena: blocking and suppletion. True ineffability (blocking) is difficult to identify in syntactic phenomena, due to the availability of paraphrased (though not systematic) alternatives. Suppletion also has not been as readily identified as a response to cross-word-boundary phonological conditioning; however, Chapter 3 presents a new study of name pair formation in English that suggests that such an effect—that is, the replacement of one form with another—to be possible under phonological conditions.

Finally, another quality that phonologically-conditioned syntactic phenomena share with phonologically-conditioned morphological phenomena is their relative rarity in their respective domains. While morphologically-conditioned phonology is commonplace, the opposite effect—the influence of phonology on morphology—is noted recurrently as relatively rare in comparison (see e.g. Carstairs-McCarthy 1998; Inkelas 2014). The same is true for phonological effects on syntactic material across word boundaries: these effects are notably rare, and when they do exist, their effects are significantly weaker compared to word-internal or morpheme-internal phonology, or to syntactically-conditioned phonological phenomena. Such decay across word boundaries of phonological constraint strength is in fact expected, though, and should be typical of such phenomena. The strength of phonological conditions has been shown to decay gradiently across boundaries (e.g., Martin 2005; 2011; McPherson & Hayes 2013); therefore, we should expect that their strength of application further decay across word boundaries. Moreover, it will be shown in the subsequent chapters that phonological conditions such as rhythm are involved in complex interactions with semantic, usage-based, and higher-order factors. That phonology must compete with other, non-phonological conditions on syntactic outputs goes towards explaining why phonological effects may be dampened in operations outside the word. What is crucial here, though, is that, even with dampened effects, the empirical behavior of phonologically-conditioned syntactic phenomena is typologically similar to more familiar effects of phonologically-conditioned morphology.

Chapter 3. Case Study 1: Rhythm in personal name choice

1 Introduction

This chapter examines syntactically-restricted data—specifically, first and last name pairs in English—to investigate unambiguously phonological effects in word choice in syntagmatic contexts. The phenomenon of word choice and its phonological conditions are taken to be a word-external parallel to word-internal, phonologically-conditioned suppletive allomorphy. Though there are not systematic alternatives as with regularized suppletion, what is shown here is that alternative lexemes are chosen as a consequence of phonological well-formedness constraints, just as alternative suppletive morphemes can be used in response to phonological conditions.

The study of phonological influences in complex syntactic constructions is often plagued by the numerous stronger semantic, syntactic, sociolinguistic, and other extra-phonological confounds that must be controlled for and potentially mask the subtle phonological effects that are present. One method of ameliorating the problem of dealing with extra-phonological semantic and syntactic factors is to explore the effect of rhythm and segmental phonology in a phenomenon of linguistic choice that is not subject to as many of the same confounds as more complex constructions are. The formation of forename-surname personal name pairs represents one such set of data.

Name pairs constitute phrases. When pronounced together, forename-surname pairs exhibit phrasal stress on the final word: for example, *Jóhn* and *Smíth* becomes *Jòhn Smíth*, with secondary stress on the first element *John* and main phrasal stress on *Smith*. To compare, genitive constructions, along with other phrases, also feature the same phrasal stress pattern: *cár* and *whéel* becomes *càr's whéel* and not **cárs whèel*. Conversely, forename-surname pairs are unlike lexical compounds in that they do not feature first-element main stress typical of compounds: *hóusewìfe* (**hòusewífe*) and **Jóhn Smìth*.

In addition to being a phrasal unit, personal names are analogous to the more complex syntactic constructions explored in later chapters because they represent an opportunity for linguistic choice. While surnames are usually fixed, the choice of an accompanying forename is free and conditioned by numerous factors, similar to, for example, the choice of the direct object or recipient immediately following a dative verb. A point of difference lies in the fact that there is a nearly infinite range of possibilities for forenames rather than a binary choice between two alternate constructions, as in the dative or genitive alternations, but this freedom of choice should not pose any significant problems for the present study.

Because personal name choices involve a limited set of constituents—content words, and specifically, names—, they are free from many of the semantic, syntactic, and otherwise extra-phonological confounds that are present in studies of more complex constructions and commonly interact with or mask rhythmic factors. There are, for example, no animacy effects to consider, since both forename and surname refer to animate entities. Furthermore, each constituent of personal names consists of only one word. In this way, the choice of personal names is similar to other phenomena, like noun-noun coordination (e.g., *peas and carrots* ~ *carrots and peas*, *Fred and Ginger* ~ *Ginger and Fred*), that remain somewhat resistant to certain sets of extra-phonological factors (Wright et al. 2005; Benor & Levy 2006; a.o.), and, as such, make for useful cases in which to test the hypothesis that the preference for rhythmic regularity between lexical stresses can influence choices made by the speaker in phrasal units.

Most of the onomastic corpus work done on English personal name choice has treated forenames and surnames individually rather than as an integrated unit, examining distributions of forenames without accounting for accompanying surnames. Previous work has also concentrated predominantly on external factors for naming practices. The ethnic, cultural, linguistic, religious, socioeconomic, and educational backgrounds and communities of the parents who choose the names for their children have been demonstrated to factor into the determination of forename choice, as shown in studies of English, Dutch, and Spanish name databases (Bloothooft & Groot 2008; Mateos & Tucker 2008; Bloothooft & Onland 2011; a.o.). Other work has pointed to effects of naming trends, popularity, and frequency in the choice of names (Tucker 2001; a.o.), as well as preferences for sound symbolism (Whissell 2001).

In the popular realm, however, a wealth of advice for baby-naming targets and highlights the relationship between forename and surname as a unit in which phonological and rhythmic well-formedness considerations are paramount. An advice columnist on an internet forum for mothers writes the following in a list of rules for baby naming:

> The baby first name's rhythm should match the last name. … Say the first, middle, and last name several times to test the rhythm. Say the first and last name together, too. (www.circleofmoms.com)

Many recommendations of the same type stressing rhythmic well-formedness in naming practices are found (e.g., Wilen & Wilen 1993), demonstrating an intuition that rhythmic preferences should in part drive the choice of accompanying forename to given surname. Herein lies a further difference between name pair formation and the alternation between genitive or dative constructions: name pairs are usually far more carefully considered than on-the-fly utterances, and, as such, we should expect to find more robust phonological effects if they do exist since they are much more consciously salient.

Based on the Principle of Rhythmic Alternation, as discussed in Chapter 1, the hypothesis for naming practices is that forenames are chosen for optimal rhythmic patterning when coupled with the given surname, all else being equal. For example, given the last name *Smith*, the forename *Súsan*, with stress on the initial syllable, should be preferable to *Suzánne*, with stress on the final syllable, because *Suzánne Smíth* would create stress clash between the first and last names while *Súsan Smíth* would form an optimal binary alternation of stress between the two elements. Similarly, a name like *Rebécca Smíth* would be preferred to *Tíffany Smíth*, because the latter forms stress lapse, with two unstressed syllables between stress peaks.

This chapter examines the rhythmic patterns of forename-surname pairs in English using a large-scale corpus of English full names culled from an online social network. Three corpus studies were performed to investigate the effect of rhythm and other phonological factors on forename choice. The results suggest that forename-surname pairs that conform to rhythmic and phonological well-formedness preferences, including the Principle of Rhythmic Alternation, are significantly more frequent than name pairs that do not exhibit the same phonological traits; these results are consistent across corpus-internal investigations and in comparison to the space of possible forename-surname combinations altogether. Furthermore, the other phonological factors tested here with rhythm are ones that have been shown to also affect syntactic choices like the genitive alternation in English, demonstrating a parallelism between the lexical choice of names discussed in the current chapter and the syntactic processes of word order and construction choice (Chapter 4).

This chapter is organized as follows. Sections 2 and 3 introduce the data used for the current study and how the predictors included in modeling and analysis were annotated and operationalized in the corpus. Section 4 details three corpus studies on rhythmic influences in personal name choice, and §5 concludes with a discussion of the results and analysis.

2 Data

In the past, large-scale investigations of full name naming practices have been rare due to data limitations and privacy concerns. Large publicly available and digitized population databases are uncommon, particularly after the de-popularization of printed phone books and the privatization of online people-search websites (Tucker 2001; a.o.). The U.S. Census releases forename-surname pairs only after 100 years have passed due to privacy protections, making any study of entire personal names based on the official census trail behind modern and current linguistic trends and patterns.

For the purposes of this study, I built a corpus of English name pairs based on the Bowes (2010) list of names from all publicly available and searchable profiles on Facebook, a popular social media website (www.facebook.com). The list was originally collected by Bowes using an automatic web crawling program, and consists of 171 million names total, with 100 million unique names. Because all of the names come from publicly searchable, voluntary users on Facebook, concerns over privacy are mitigated; to provide for further anonymity, associations between each name and its corresponding profile URL were not maintained in the corpus used herein.

The Bowes (2010) list then underwent processing via a combination of automatic and hand filtering in an effort to exclude undesirable data, including but not limited to names of businesses (e.g., *Rainforest Café*), non-English names (e.g., *Rajesh*), and obvious nicknames, aliases, and fictional characters (e.g., *Lord Voldemort*).[26] First and last name pairs that occurred only once in the list were excluded as a rough cut through the data: while this omitted a small portion of legitimate unique names, it also provided an efficient way to eliminate non-legitimate names. Another rough cut was made by excluding names that were not present in the Unisyn lexicon (Fitt 2001), which helped to ensure that names were English and not of obviously foreign origin. Because the subject of this investigation is forename-surname pairings, full names with more than two orthographical words were excluded, as were name pairs in which the first or last name contained only one orthographic letter. This eliminated the possibility of double first names or middle names, but evened the playing field to those who identified with only first and last name information.[27] In sum, the final corpus used for the current studies consists of 41 million forename-surname pairs total, of which there are 3.3 million unique name pairs. Despite the filtering efforts, there no doubt remains some acceptable amount of noise in a dataset of this size.

3 Predictors

The corpus was first annotated with lexical stress, segmental, and syllable information using automatic annotation based on the American English Unisyn lexicon (Fitt 2001) and manual annotation. Several factors were operationalized and included in the anal-

[26] Although, we would expect that such names would also follow the Principle of Rhythmic Alternation.

[27] It should be noted that middle names could also play into rhythmic considerations.

ysis: rhythm, alliteration, avoidance of adjacent segmental identity, avoidance of adjacent sibilancy, avoidance of rhyme, individual forename and surname frequencies, and forename popularity. In addition to the predictors detailed in the following sections, there are other potential determinants of personal name choice that were not controlled or accounted for due to the nature of the data. In particular, person-specific information was unavailable, including familial preferences given naming practices between generations, as well as other socioeconomic, geographic, cultural, linguistic, and gender information.

3.1 Rhythm

To operationalize the Principle of Rhythmic Alternation, I follow the *Eurhythmy Distance* (ED) measure developed by Shih et al. (to appear). Eurhythmy Distance is a measure of how far from perfectly alternating binary rhythm a given inter-stress interval is, and is calculated by taking the absolute value of the number of unstressed syllables between stress peaks minus one, as formulated in (28).

(28) ED = | # of unstressed syllables between stresses – 1 |

In the eurhythmy distance measure, a ED of 0 indicates that the construction—in this case, the forename-surname pair—perfectly satisfies binary rhythm, with exactly one unstressed syllable between two stress peaks at the border of the first and last names, as demonstrated in (29).

(29) *Súsan Smíth*

$\underbrace{\text{S W S}}_{1}$ ED = | 1 – 1 | = 0

An ED value greater than 0 indicates that the construction is not optimally rhythmic:

(30) a. *Suzánne Smíth*

$\mathrm{W}\ \underbrace{\mathrm{S}\qquad \mathrm{S}}_{0}$

ED = | 0 – 1 | = 1

b. *Mélanie Fitzgérald*

$\underbrace{\mathrm{S\ WW\ W}\quad \mathrm{S}}_{3}$

ED = | 3 – 1 | = 2

In (30a), the name pair *Suzanne Smith* has clashing stresses across the forename-surname border and is one step away from eurhythmic alternation, as reflected by its ED score of 1. Example (30b) demonstrates that a name pair with lapse between its two components, as in *Melanie Fitzgerald*, incurs an ED score of 2. The greater the Eurhythmy Distance value, the farther from perfectly alternating rhythm a given name is.

The hypothesis tested here is that the likelihood of a forename-surname pair will depend on its rhythmic structure: the greater the Eurhythmy Distance score, the less likely the forename-surname pair should occur. To take the examples from above, *Susan Smith* has 2172 occurrences in the facebook corpus, as compared to 550 occurrences of the clashing *Suzanne Smith* and 27 occurrences of the lapsing *Melanie Fitzgerald*. Overall, name pairs with perfectly alternating binary rhythm at the forename-surname boundary make up the majority of the corpus (54.69%; n = 22,498,578), occurring significantly more often (χ^2 = 19954933.851, p < 0.0001) than name pairs with clashes or two syllable lapses (44.25% of the total corpus; n = 18, 204,796), and name pairs with trisyllabic lapses (1.06% of the total corpus; n = 435,905).

3.2 Other predictors

ADJACENT SIBILANCY. Observations of genitive construction choice have shown that English speakers tend to avoid sequences of immediately adjacent sibilants, including [s], [z], [ʃ], [tʃ], [ʒ], and [dʒ] (Menn & MacWhinney 1984; Zwicky 1987; Hinrichs & Szmrecsányi 2007; et seq.). Morphological –*s* genitive constructions where an –*s* possessive morpheme will occur next to a final sibilant in the possessor—for example, *the judge* + -*'s* + *descendants*—are far less likely to occur, even though phonological processes such as haplology and [ə] epenthesis exist to repair the sibilant adjacency. Sibilant adjacency for the present corpus was automatically coded using the Unisyn lexicon and the Unisyn supplemental lexicon in order to control for and investigate the effect of adjacent sibilant avoidance on name choice formation, with the expectation that forename and surname pairs that feature adjacent sibilant sounds, as in *Josh Smith*, *Charles Smith*, or *Charles Jones*, will be more frequently avoided.[28]

ADJACENT PHONOLOGICAL IDENTITY (OCP). Similarly to the avoidance of adjacent sibilant sounds, adjacent identical segments are often avoided cross-linguistically, in a general Obligatory Contour Principle-type (henceforth, OCP) ban (Leben 1973; Goldsmith 1976). Identical segments across forename and surname boundaries (e.g., *Sam Madison*) were coded automatically. Vowels were treated as identical with all other vowels (e.g., *Tami Erikson*); consonants were marked as identical only when they are the same segment as their adjacent counterparts.

ALLITERATION. Alliteration—the repetition of identical initial segments—is a common poetic device in artistic forms as well as in idioms and common phrases in English (e.g., *bad to the bone*). Boers & Lindstromberg (2005) found through a series of recall experiments that recall in English is significantly improved when the tokens involved alliterating sounds. Given this bias in English for alliterative segments, it is possible that alliterative names are also preferred, for a variety of artistic and perhaps processing and memory based factors. For alliteration, identical initial segments of forenames and surnames were coded automatically. Vowels were treated as identical with all other vowels (e.g., *Elizabeth Erikson*); consonants were marked identical only when the initial segments of the names were identical.

[28] As a note for future research, it is possible that complete segmental identity is more preferred than near segmental identity: for example, *Russ Smith* ([s]-[s]) preferred to *Rush Smith* ([ʃ]-[s]). Complete identity results in a geminate, which is a potentially more stable configuration than segments that are too discomfortably similar to coexist (see e.g., Inkelas & Shih 2013 for an overview of such effects).

RHYME. In the popular baby-naming literature, it is generally advised that rhyming first and last names should be avoided. Thus, a rhyming name such as *Michelle Bell*, which has 348 occurrences in the facebook corpus, should be disprefered in comparison to a name that does not rhyme, such as *Mike Bell* (n = 643). Rhyme was operationalized by automatically coding the identity of the syllable rhymes including and following the primary stressed syllable based on Unisyn lexicon pronunciations. If the segments in every syllable rhyme including and following the primary stressed syllable were identical, the names were considered rhyming. For example, *Michelle Bell* is considered a rhyming pair because the syllable rhymes including and following the stressed syllable—in this case, [ɛl]—are identical in the forename and surname. *Mitchell Bell*, however, would not be considered a rhyming pair because [ɪ - əl] of *Mitchell* is not completely identical with [ɛl] of *Bell*.

POPULARITY AND FREQUENCY. Trends in name popularity can affect the choice of first names (Tucker 2001; a.o.). To control for this tendency, the sum of the frequencies of the 400 most frequent baby names—200 male and 200 female—were taken from each decade ranging from 1950 to 2000 from the Social Security Administration list of popular baby names. These decades were chosen to represent the age range of the majority of facebook users. Frequency of the last names was controlled for corpus-internally, using the frequency of each surname in the facebook corpus.

4 Corpus studies

The following sections present three corpus studies of the effect of rhythm on name pair formation using poisson and logistic regression models. The first two studies (§§4.1–4.2) investigate the role of rhythm, in addition to the other phonological predictors, in predicting the frequency of occurrence of a given name pair. The third study (§4.3) explores the role of rhythm and phonological predictors in determining the actual occurrence of a name pair. In the following models, all predictors were centered by subtracting their means, and all numerical predictors were also divided by twice their standard deviations. Centering and standardizing the predictors allows for direct comparison of the model coefficients, and protects against harmful effects of collinearity (Gelman 2008). The independent frequency and popularity predictors were also logarithmically transformed to remove skew caused by outliers in the frequency distributions.[29]

4.1 Study 1: Frequency of name pairs in Facebook corpus

The first study presented here examines rhythm in frequent name pairs. Under the Principle of Rhythmic Alternation, the prediction is that more frequent personal name pairs are more likely to follow rhythmic well-formedness preferences, as well as other phonological preferences, because speakers should gravitate towards using name pairs that are phonologically well-formed, all else being equal. *Sarah Smith* (*n* = 5039), for example, should be a highly frequent name, because it features alliteration, no OCP violations, no rhyme, and perfectly alternating binary stress across the forename-surname boundary. Names such as *Donovan Ladd* (*n* = 2) and *Dorcus Scott* (*n* = 2) should be much less frequent, due to rhythmic and OCP violations, respectively.

A subset of the facebook corpus was used for this study, consisting of all polysyllabic forename and monosyllabic surname pairs (e.g., *Sarah Smith*) (*N* = 806,233). Monosyllabic surnames were used in order to test the pairing of a chosen forename with surnames of a single rhythmic structure. Name pairs with monosyllabic forenames were excluded from the study because a pair of monosyllabic first and last names would automatically result in a stress clash. The study presented in §4.2 deals with the choice between polysyllabic and monosyllabic forenames.

---

[29] First name Social Security Administration census measures were logarithmically transformed by adding 1 before taking the log due to the zero observations for the names that did not occur in the census: = log(Census Popularity + 1).

The model results of the study are provided in Table 1 below. Quasi-poisson modeling was used for Studies 1 and 2 to adjust for over-dispersion in the data.

Table 1. Quasi-poisson regression estimates (Study 1)[30]

Factor	Estimate	Std. Error	Wald Z	Pr (>|Z|)	
Intercept	1.61255	0.0042	385.454	<0.0001	***
Eurhythmy distance	-0.0987	0.0057	-17.328	<0.0001	***
Last name frequency	1.1091	0.0021	529.730	<0.0001	***
First name popularity	1.3169	0.0076	174.162	<0.0001	***
First name freq (resid)	0.067	0.00018	377.701	<0.0001	***
OCP violation = Y	-0.0134	0.01538	-0.89	0.37363	
Adjacent sibilant = Y	-0.07812	0.02496	-3.13	0.00175	*
Alliteration = Y	0.04976	0.00983	5.062	<0.0001	***
Rhyme = Y	-0.2135	0.11947	-1.757	0.07387	
Last freq * First freq	0.12507	0.00404	30.924	<0.0001	***
N	806233		adj pseudo R^2	0.81	
quasi log likelihood	-76149.08 (df = 10)***		dispersion parameter	53.76555	

. significant at $p < 0.05$, * significant at $p < 0.01$, ** significant at $p < 0.001$, *** significant at $p < 0.0001$

There is still a fair amount of variance and variability in the residuals left unexplained by the model given in Table 1; however, this is to be expected given the external linguistic factors that were not included here, including but not limited to social and cultural considerations. Even so, we can see that, controlling for the frequency and popularity of surnames and forenames, respectively, phonological well-formedness preferences contribute to the frequency of a given name pair.

[30] Adjusted McFadden's pseudo-R^2 calculated as 1 – (Residual deviance / Null deviance). Quasi likelihood ratio calculated by dividing log likelihood of regular poisson model by the dispersion parameter for quasi-poisson model (following Lebreton et al. 1992)

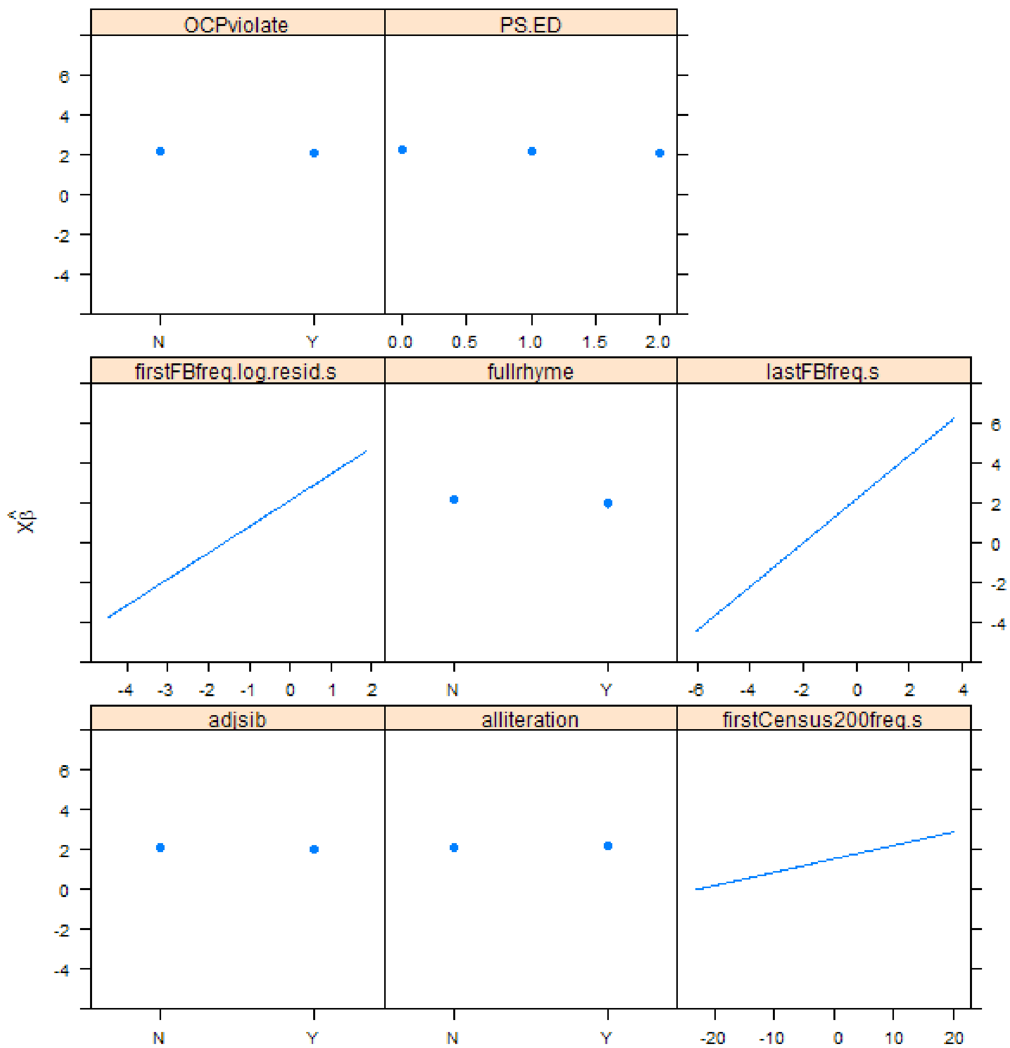

Figure 2. Partial effects of model predictors (with all other predictors held constant)

Figure 2 below illustrates the variable effects plots of each predictor, with all other predictors held constant. The increasing log frequency of the name pair in the facebook corpus is given on the *y*-axis, with the values of the predictors on the *x*-axis. Confidence intervals are included in the figure; however, given scaling and the tight fit with the data, the intervals may not be visible here. Figure 3 provides the decrease in the model's goodness-of-fit (increase in -2 log likelihood) with each predictor removed in turn.[31]

The phonological predictors all behave as expected. Adjacent sibilants and identical segments at the forename-surname boundary result in less frequent name pairs. The avoidance of adjacent sibilants is the significantly more restrictive OCP predictor. The presence of adjacent sibilants sees a 7.51% decrease (100 × (1 – exp(-0.07812)) in name pair frequency. A general OCP violation also demonstrates a decrease, but the effect is smaller and not reliable in this model. The avoidance of rhyming name pairs is a trending effect: name pairs that rhyme tend to occur less frequently than name pairs that do not rhyme, with a 19.22% decrease in frequency between non-rhyming and rhyming pairs, when all else is held constant. The presence of alliteration has the opposite result. Frequent name pairs in the facebook corpus are more likely to be alliterative. All else being equal, alliterative name pairs are 1.051 times more frequent than non-alliterative pairs.

[31] Model likelihoods calculated on poisson models.

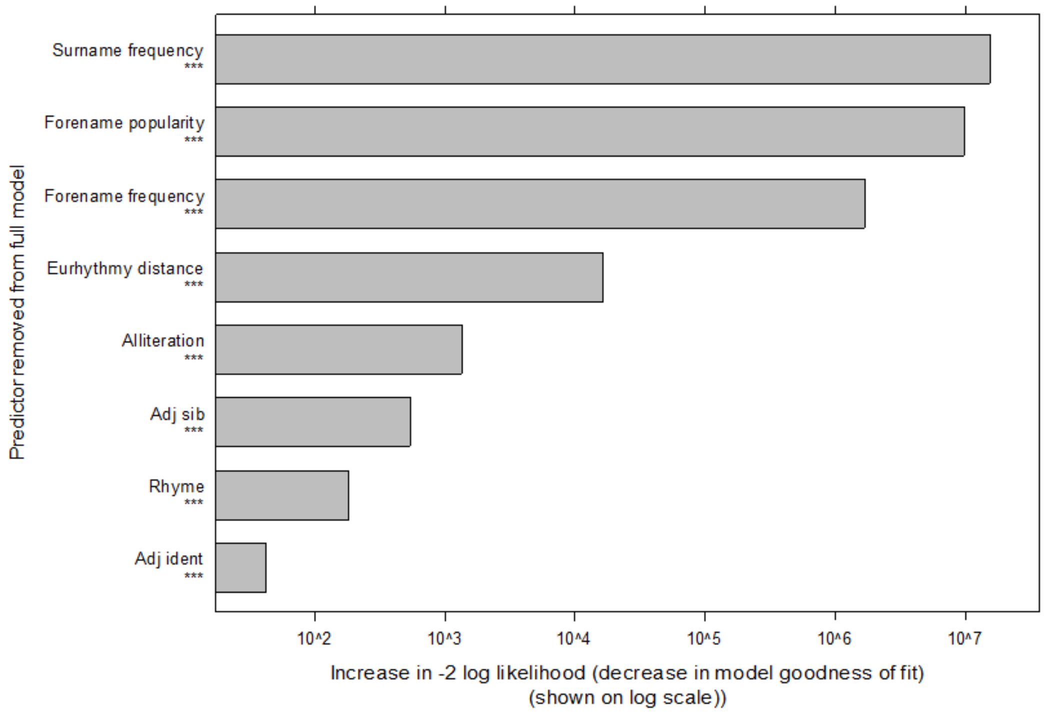

Figure 3. Increase in -2 log likelihood (decrease in model goodness-of-fit) if factor is removed.

With all other predictors held constant, the results from Study 1 also show that frequent name pairs are more likely to exhibit binary rhythm across the forename-surname border, avoiding clash and lapse between stressed syllables. The partial effects plot illustrates that as Eurhythmy Distance increases—that is, as the name pair construction moves farther away from perfectly alternating rhythm—the frequency of the name decreases. The regression coefficient reveals that for every one unit increase in the Eurhythmy Distance measure, there is a 9.4% decrease in the frequency of the given name pair. The model goodness-of-fit analysis in Figure 3 shows that rhythmic alternation preferences, amongst all of the other phonological predictors, contributes the most explanatory power to the model, second only to the frequency and popularity control variables.

The results of the first study presented in this section demonstrate that the frequency with which a name pair occurs is conditioned in part by phonological well-formedness preferences, including, in large part, by eurhythmic considerations.

4.2 Study 2: Frequency of name pairs with iambic surnames

While the results of Study 1 in §4.1 demonstrated that rhythm influences the use of forename-surname pairs, it is possible that these results are a consequence of the range of rhythmic patterning available in English names. The phonological make-up of English forenames and surnames predisposes pairs to a perfectly alternating binary rhythmic pattern. The majority of English last names, including monosyllabic and polysyllabic names, are trochaic and stress-initial (n = 3,019,456, 91.73%)—for example, *Jóhnson*, which has stress on the first syllable. Likewise, the majority of English first names are trochaic and end in an unstressed syllable (n = 1,716,069, 52.14%), as in *Dávid*. Thus, given the phonological make-up of English names, perfect rhythmic patterning is already expected.

To investigate whether there is an effect of rhythmic preferences in English above and beyond the chance combination of the majority of English forenames and surnames, the study presented in this section explores the role of rhythm in name choice with iamb-initial, polysyllabic surnames. Iamb-initial surnames in the corpus include *Buchanan, Burnett, Levine, Maloney, Marie, McDonald, Montgomery,* and *Munro*, amongst others. If rhythmic well-formedness preferences are indeed at work, the prediction is that for these iamb-initial last names, there will be a greater affinity to use stress-final first names or monosyllabic first names to create a eurhythmic strong-weak-strong pattern across the forename-surname boundary.

For the purposes of this experiment, a subset of the facebook names corpus was used, consisting of name pairs that have polysyllabic last names with peninitial stress (e.g., *Fitzgérald*) and polysyllabic or monosyllabic first names (n = 286042). As in the first study, frequency of the name pair in the facebook corpus was used as the dependent variable under the hypothesis that the likelihood of a recurring name pair depends in part on its phonological well-formedness: names that conform to phonological preferences—and in particular, rhythmic ones—will be more often preferred to less phonologically-ideal alternatives. Thus, all else being equal, *Suzánne Fitzgérald* and *Súe Fitzgérald* should be preferred to *Súsan Fitzgérald* in order to avoid lapse between the rightmost stress of the forename and the leftmost stress of the surname.

Results of the model demonstrate that Eurhythmy Distance is again reliably predictive of name pair frequency. Its effect is shown in the partial effects plot in Figure 4 below. The more frequent a forename-surname pair is, the more likely it will be rhythmically optimal (β = -0.04386, *S.E.* = 0.00589, $p < 0.0001$ ***),[32] even when the surname begins with an iamb. More frequent name pairs are significantly more likely to use monosyllabic or stress-final polysyllabic first names so as to avoid lapses in rhythmic patterning. With every one unit increase in the Eurhythmy Distance measure, the frequency of the name pair decreases by 4.29% (100 × (1 – exp(-0.0436))).

[32] A mixed poisson regression controlling for first names and last names as random effects, run on a subset of data that controls for OCP (both general and sibilant-specific) and avoidance of rhyme (n = 266674, with 6065 unique first names and 2648 unique last names), demonstrates a trending effect of rhythm only, but in the expected direction (β = -0.036733, *S.E.* = 0.021575, $p < 0.0886$).

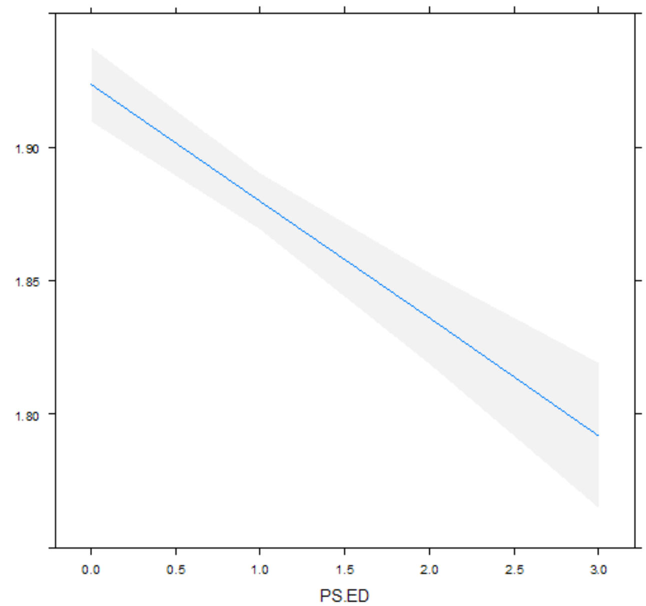

Figure 4. Partial effect of Eurhythmy Distance (PS.ED) , with all other predictors held constant

By restricting the surnames to iamb-initial stress patterns, we see from the study in this section that the rhythmic well-formedness preference for binary alternating stress plays a role in conditioning first name choice for an accompanying last name even though the majority English names are trochaic and thus already predisposed to rhythmic alternation. We find here that rhythmic names are more likely to occur more frequently: in particular, iamb-initial last names that are paired with monosyllabic or stress-final polysyllabic first names will occur with greater frequency.

4.3 Study 3: Attested versus unattested personal names

The results of Studies 1 and 2 (§§4.1–4.2) demonstrate that name pairs that abide by rhythmic and phonological well-formedness preferences are more common than ones that violate well-formedness considerations. The study in this section examines the influence of rhythm and other phonological factors in determining the names people will choose, given the range of possible first and last name pairs. If the Principle of Rhythmic Alternation conditions our choices in speech, name pairs that occur regardless of frequency should exhibit more rhythmic patterning than name pairs that never occur. A name such as *Deníse Fitzgérald*, which has only thirty-eight total occurrences in the facebook names corpus and features binary alternating rhythm across the forename and surname boundary, should still be better than a name such as *Ánalise Fitzgérald*, which is unattested in the facebook names corpus and has a long lapse between stressed syllables. Under the hypothesis here, no matter how innovative or original a name is, the name should still strive for perfectly alternating binary rhythm between forename and surname.

Using only polysyllabic first and last names, a parallel dataset of name pairs that are unattested in the facebook names corpus was generated. Forenames and surnames from the corpus were randomly shuffled, and the random pairs were then checked against the attested name pairs. Name pairs that were already attested in the corpus were removed, and only generated name pairs that did not occur in the corpus were retained. This generated dataset provides a baseline with which to compare attested name pairs for phonological well-formedness. All told, the dataset of real and generated polysyllabic forename-surname pairs together contains 3,461,906 name pairs, with 1,649,342 attested name pairs and 1,812,564 randomly generated 'fake' name pairs.

The same logistic regression model, predicting real versus generated name pairs as the dependent variable, was fit over ten random subsamples of 500,000 tokens each and also tested on five random subsamples of 1,000,000 tokens each. The results showed very little to no variation across subsamples, and a representative model of a random subsample of 500,000 tokens is reported on here, due to big data limitations. The final model was also verified over a random subsample of 500,000 tokens using 1000 runs of bootstrap resampling with less than 0.0001 optimism in the R^2 and *Dxy* statistics. Table 2 below provides the results of the model. Figure 5 illustrates the decrease in model goodness-of-fit (increase in -2 log likelihood) with each predictor removed in turn. The partial effects plots are given in Figure 6, with all other predictors held constant. A greater log odds value on the *y*-axis indicates a greater likelihood of a real, attested name pair, and a smaller log odds value indicates a greater likelihood of a fake, generated name pair.

Table 2. Logistic regression estimates, on N = 500,000 subsample

Factor	Estimate	Std. Error	z value	Pr (>|t|)	
Intercept	-0.381	0.006	-60.13	<0.0001	***
Eurhythmy dist	-0.945	0.039	-24.53	<0.0001	***
Last name freq	0.923	0.003	291.62	<0.0001	***
First name pop	0.131	0.0004	297.85	<0.0001	***
OCP = Y	1.085	0.091	11.99	<0.0001	***
Adj sibilant = Y	-7.554	0.468	-16.14	<0.0001	***
Allit = Y	5.324	0.161	33.04	<0.0001	***
Rhyme = Y	85.019	4.134	20.57	<0.0001	***
N	500000		adjusted R^2		0.819
model χ^2	475797.64 (df = 7) ***		%correct (%baseline)		97.1 (91.1)
adjusted *Dxy*	0.943				

. significant at $p < 0.05$, * significant at $p < 0.01$, ** significant at $p < 0.001$, *** significant at $p < 0.0001$

With the exceptions of rhyme and general adjacent identical segment avoidance, the phonological predictors behave as expected. Attested name pairs are more likely to alliterate than unattested name pairs: alliteration between the forename and surname significantly increases the chances of an actual name pair, all else being equal. There is also an avoidance of adjacent sibilant segments in attested name pairs.

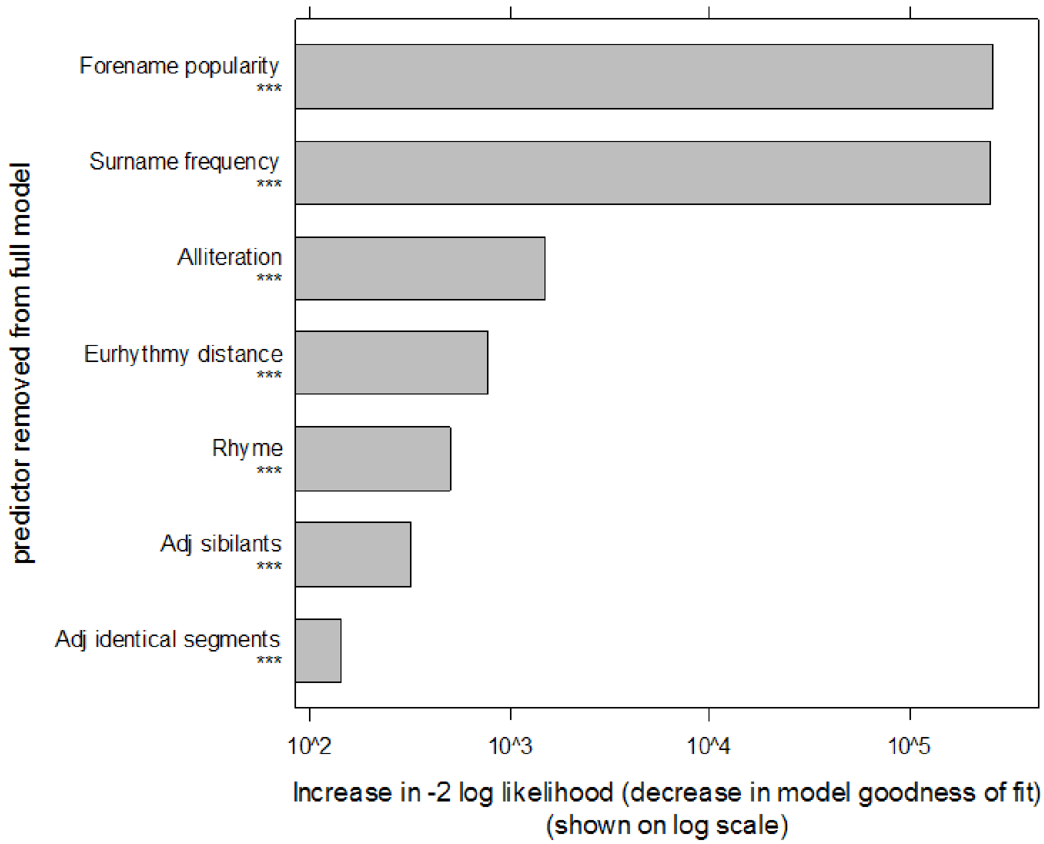

Figure 5. Increase in -2 log likelihood (decrease in model goodness-of-fit) if factor is removed

The behavior of rhyme avoidance and general adjacent identical segment avoidance are the opposite of the expected pattern. The model results indicate that actual name pairs in the facebook names corpus are more likely to rhyme and have OCP violations than name pairs not found in the facebook names corpus. It is unclear how to interpret these results. One explanation is the nature of the data: it is possible that the names provided by the corpus are a priori not very likely to rhyme and, similarly, are not very likely to co-occur with identical segments at the forename-surname border.

The effect of Eurhythmy Distance is clear. For roughly every one unit increase of the Eurhythmy Distance measure (two standard deviations = 1.09), the likelihood that a given name pair is a real pair from the facebook corpus decreases by 38.9% (odds ratio = 0.389). That is, the farther away from perfectly alternating rhythm a given name pair is, the more likely the name pair is a randomly generated, ‘fake’ token.

As Figure 5 shows and as we saw in the previous two studies, rhythmic well-formedness is again one of the phonological predictors that contributes the most explanatory power to the model, second only in this instance to alliteration.

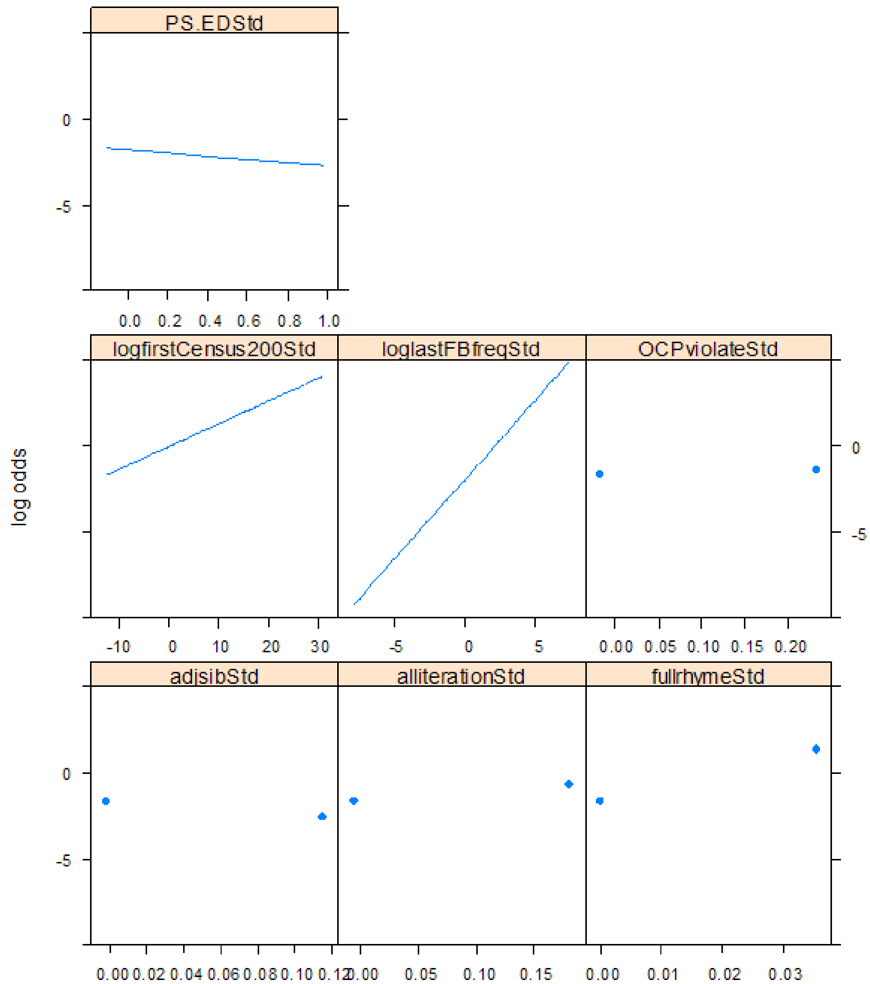

Figure 6. Partial effects of model predictors (with all other predictors held constant)

5 Discussion

Controlling for available non-phonological factors, the results from the three studies presented above demonstrate that phonological preferences affect personal name choice, both in terms of how frequent a given name pair is and the chances with which a name pair will be chosen for use by speakers. In particular, rhythmic well-formedness preferences, the avoidance of adjacent sibilants, and alliteration are robustly reliable predictors of frequent and attested name pairs. To a somewhat lesser extent, the avoidance of rhyming forenames and surnames and the avoidance of adjacent identical segments at the first and last name boundary also contribute to determining name choices. These phonological predictors are discussed in turn.

Of most interest here is the effect of rhythm on English personal name choice. In all of the three studies in §4, rhythmic regularity was a significant and important predictor of both name pair frequency and use. The farther away from perfectly alternating binary rhythm a forename-surname pair is, the less frequent the pair will be, and the greater the chances that the pair will not be used by speakers as an actual name. Amongst the phonological predictors, rhythm regularly contributed the most explanatory power in the models, indicating that rhythmic alternation plays a large role in speakers' considerations of name pair choice.

Similarly, the presence of alliterative initial segments and the avoidance of adjacent sibilants were significant and reliable predictors across all three of the studies. Frequent name pairs are more likely to be alliterative. Speakers also tend to avoid sequences of first and last names that have sibilants at the boundary of the two names. Crucially, both rhythm and the avoidance of adjacent sibilants are factors that have been identified to influence genitive construction choice (Shih et al., to appear) (see also see Chapter 4), in addition to other word order choices such as binomial pair ordering. These parallels suggest the same phonological preferences are active across many linguistic processes, including personal name choice; therefore, the effects that we uncover in one process warrants testing for generalizability to other construction types.

The results of Study 1 (§4.1) also suggest that the avoidance of adjacent identical segments in part influences name choice, much as the avoidance of adjacent sibilants does. This OCP-type effect was not as robust of a predictor as the ban on adjacent sibilants in the studies presented above. It is possible that the OCP definition as tested here was not correctly tuned. That is, the definition of identical adjacent segments used here does not distinguish between possible biases to avoid certain clusters per English phonotactic preferences, nor does it identify avoidances of classes of similar sounds, like the sibilant ban does. For example, [l] – [r] clusters are likely to be more dispreferred, as in *Carl Rogers* (n = 104), than [m] – [m] sequences, as in *Tom Mitchell* (n = 368). Martin (2007) reported this type of gradient OCP avoidance within English words and baby names—and particularly with [l] – [r] sequences, and it is likely that such an effect holds across word boundaries in a tight-knit phrase like the first and last name pair.

The avoidance of rhyming names is another factor that contributes to personal name choice, though not as strongly and reliably as the other phonological predictors. Like the predictor for adjacent identical segments, it is possible that the definition of rhyme tested here is not the optimal one and fails to accurately capture the aversion to rhyming pairs that English speakers might have. Speakers could be more liberal in their tolerance of rhyme, allowing anything but the strictest rhyming pair: for instance, *Michelle Bell* (n = 348) would be tolerated, but *Mel Bell* (n = 39) would not. This effect is left for closer study in the future.

There are yet more factors of forename and surname pairing that remain unexplored in this chapter, in addition to various social, cultural, and historical considerations. Gender, for example, is likely to play a significant role in rhythmic preferences, given the rhythmic differences in structure between male and female first names (Wright et al. 2005). While alliteration is a reliable predictor of name pair choice and frequency, orthographic alliteration might also influence naming, especially if a pair of repeating orthographic letters form desirable initials. The alliteration factor considered here only included alliteration when the names are pronounced; in orthographic alliteration, a name like *Sharon Smith* would be considered alliterative, even if the initial phonemes [ʃ] and [s] do not match.

Another possible phonological conditioning factor of name pairing might also be phonotactic and syllable structure preferences. In their study of binomial name ordering (e.g., *Fred and Ginger ~ Ginger and Fred*), Wright et al. (2005) suggested that the pressure to have onsets (the ONSET constraint) and the dispreference of codas (NO-CODA) contributes to the ordering of names in a pair. For example, given the names *Mike* and *Anna*, *Anna* would be preferred in the second position because the [d] of the conjunction *and* could serve as an onset to *Anna* and because *Mike* provides a [k] onset for *and*. Such syllable constraints might play a role in name pair choice; for instance, *Ann Adams* (n = 183) would be a preferable name pair in comparison to *Annie Adams* (n = 74) because the final [n] in *Ann* can act as an onset to *Adams*.

Finally, recent work by Ramscar et al. (2013) suggests possible information-theoretic motivations for name choice that were not considered in this case study, namely that in naming practices, speakers seek to minimize uncertainty and optimize information sequentially, from first name to last. The predictors of popularity and frequency used here control for some of the possible information-theoretic effects; however, more study is reserved for future work.

6 Conclusion

Names are an intermediary form between artistic language forms and productive speech. They represent a tightly-knit construction fairly independent of the influences of semantic or pragmatic factors found in other constructions—for example, animacy or givenness factors that affect genitive construction choice (see discussion in Chapter 4). The results presented here, from a large-scale, corpus-based case study of English personal name choice, demonstrate that naming practices are subjected to phonological well-formedness conditions, just as other linguistic processes are. Rhythmic regularity, the avoidance of adjacent identical segments, the avoidance of rhyme, and the affinity for alliteration all contribute in some way towards the realization of a forename-surname pair in English. Many of these predictors—in particular, rhythm and the homophony-avoidance constraints—are known to be active elsewhere in the language as well as cross-linguistically (e.g., Golston 1995). The study of name pairs, therefore, is a valuable testing ground for investigating these types of effects that may also appear more generally throughout language use in other construction types.

Chapter 4. Case study 2: Rhythm in genitive construction choice

1 Introduction

This chapter[33] presents a study of the role of rhythm in conditioning morphosyntactic alternation. The previous chapter demonstrated that rhythmic well-formedness preferences, along with other phonological conditions, can be met via word order choices in syntagmatic contexts. The goal of this chapter is to examine the extent to which rhythmic well-formedness conditions can be satisfied via morphosyntactic operations such as word order rearrangement and alternative construction choice. The expectation under the Principle of Rhythmic Alternation (see discussion in Chapter 1) is that syntactic choices will also be subject to the optimization of rhythmic patterns.

English has two syntactically distinct constructions for expressing the possessor and possessum relationship: the *s*-genitive and the *of*-genitive:

(31) a. the car's wheel b. the wheel of the car

The *s*-genitive (31a) is a single noun phrase, where the possessor *car* occurs before the possessum *wheel* accompanied by the possessive clitic *–s*. The *of*-genitive (31b) consists of two noun phrases, with the possessor *car* located in a prepositional phrase headed by *of*. Insofar as cliticization is considered a morphological process, the alternation between genitive constructions in English can be viewed as an almost-

[33] This chapter is based on joint work with Jason Grafmiller, Richard Futrell, and Joan Bresnan. A condensed version of this material appears as Shih, Grafmiller, Futrell, & Bresnan (to appear). The results herein have been updated since the Shih et al. (to appear) version (e.g., using more up-to-date control variables: see §3.2), but the results themselves—in particular, the results of rhythmic conditioning—remain largely unchanged. The first author contributed in the original data collection and annotation, provided the phonological analysis and interpretation, and was responsible for the original draft of the work. Moreover, Shih is solely responsible for any updates since the Shih et al. (to appear) version. All remaining shortcomings herein should be attributed to the first author.

systematic periphrastic relationship between a morphological construction using the clitic *–s* and an analytic construction using *of*. The goal here is to see whether this type of alternation that exists at a word-external level can be phonologically-conditioned.

The choice between the two English genitive constructions is not a free one. Rather, the choice of one genitive construction over the other has been shown to be conditioned by the interaction of numerous semantic, syntactic, and sociolinguistic factors (e.g., Rosenbach 2002; Hinrichs & Szmrecsányi 2007; Kreyer 2003; Szmrecsányi & Hinrichs 2008; Tagliamonte & Jarmasz 2008). Phonological factors including homophony avoidance of adjacent sibilant sounds caused by the *–s* clitic (e.g., *the bu**s's** wheel*) have also been demonstrated to be robust factors in genitive construction choice (Zwicky 1987; Hinrichs & Szmrecsányi 2007; Szmrecsányi & Hinrichs 2008; a.o.).

In this study, we examine the influence of rhythm in predicting genitive construction choice in spoken English. We do so by incorporating rhythmic factors into a single model of genitive choice alongside previously identified predictors using logistic regression modeling. Three measures of rhythmicity are tested: [1] Eurhythmy distance, as described in Chapter 1; [2] *CLASH and *LAPSE; and [3] Comparative eurhythmy distance, which factors in both the rhythmic patterning of the *s*-genitive and *of*-genitive constructions. We find that while rhythm—as measured by Eurhythmy distance—significantly influences construction choice, its explanatory role is small relative to other known predictors. Thus, rhythm—and phonological factors at large—must not be discounted in studies of syntactic variation, but the converse is also crucially true: rhythm alone does not do or explain everything.

This chapter is organized as follows. Sections 2 and 3 present our spoken English genitive data and introduce each of the predictors in our model, respectively. Results of the analysis based on the Eurhythmy distance measure are in §4. Section 5 discusses the results and presents comparisons with the other measures of rhythm. Section 6 concludes.

2 Data

Our study utilized spoken data from the manually parsed Penn Treebank portion (Marcus et al. 1993) of the Switchboard corpus of American English (Godfrey & McDaniel 1992) under the hypothesis that rhythmic and phonological effects will be most apparent in spoken contexts. Exploration of rhythm in written data is saved for further research (see Grafmiller forthcoming). The Switchboard corpus consists of telephone conversations between native American English speakers who did not know each other and were assigned random, predetermined conversation topics.

The key criterion for identifying the data in this study was the reversibility and interchangeability of the *s*-and *of*- genitive constructions. Following previous work on genitive construction choice (Rosenbach 2002; Kreyer 2003; Hinrichs & Szmrecsányi 2007; Szmrecsányi & Hinrichs 2008; a.o.), we only included constructions whose alternatives were equivalent and possible paraphrases: e.g., *the doctor's patients ≃ the patients of the doctor*. Excluded, then, were constructions where the *s*- and *of*- alternatives were not interchangeable, all of which have been previously identified and include the following (Quirk et al. 1985; Biber et al. 1999; Rosenbach 2002; 2005; Kreyer 2003):

- Post-genitives: *We meet at Bill's* ≠ **We meet at of Bill.*
- Genitives without noun heads: *the cost of providing the startup* ≠ **providing the startup's cost*
- Quantitative constructions: *a cup of soup* ≠ *a soup's cup*
- Qualitative constructions: *this kind of work* ≠ **this work's kind*
- Material constructions: *a crown of gold* ≠ **gold's crown*
- *Of*-constructions with premodifying quantifiers: *most of the people* ≠ **the people's most*
- Descriptive genitives: *women's magazines* ≠ *the magazines of {the|some} women*
- Indefinite possessums: *a book of a teacher* ≠ *a teacher's book*

- Fixed expressions: *arm's reach ≠ the reach of the arm*

Additionally, Rosenbach (2002) notes that pronominal possessors appear nearly categorically in the *s*-genitive form; thus, for the purposes of this study, we did not consider genitives with pronominal possessor or possessum NPs, following previous work (e.g., Hinrichs & Szmrecsányi 2007).

Genitives were chosen from the Treebank Switchboard corpus using a combination of automatic Tgrep2 filtering and manual coding. The four researchers collaborating on this study each coded a portion of the corpus, excluding constructions listed above, and cross-checked their results with the others. Our data was then checked once more for consistency by the second author. Animacy information for each noun was derived from the LINK annotations of the corpus (Zaenen et al. 2004), and demographic information about the speaker of each utterance was extracted using perl scripts from Jaeger (2005). We concluded with 1124 genitives, of which a few more had to be excluded due to missing or incomplete contextual information from Switchboard. In sum, the corpus has 1107 genitives, with 653 instances of *of*-genitives (59%) and 454 instances of *s*-genitives (41%).

3 Predictors

This section presents the conditioning factors coded in our data.

3.1 Rhythm

Before being able to examine rhythm in the genitive alternation, we first annotated our dataset with lexical stress information using automatic annotation of both primary and secondary stress based on the Carnegie Mellon University Pronouncing Dictionary (Weide 1993). Since we are interested in the simple alternation between stressed and unstressed syllables, we chose to collapse the distinction between primary and secondary stress; thus, both primary and secondary stressed syllables are, for our purposes, considered stressed syllables, forming a binary distinction between syllables that are stressed and those that are not. Words that were not found in CMU were manually coded by the first author for lexical stress and syllabification, following CMU annotations as closely as possible. Using CMU as the source of our lexical stress annotations provides us with a way to approximate speakers' stored lexical information about a word's phonological properties—in particular, stress—independent of other phonetic and syntactic pressures and effects during the speech act. A study of stress patterns based on the phonetic stream in the Switchboard conversations is left to future research. The stressed annotations from CMU were randomly hand-checked for accuracy.

We hypothesize that the Principle of Rhythmic Alternation influences the choice of genitive constructions in English. All else being equal, given a pair of possessor and possessum NPs, speakers should, under our hypothesis, choose the more eurhythmic construction, be it the *s*-genitive or the *of*-genitive. Take, for example, the possessor-possessum pair in (32): *the children* and *the voices*.

(32)	a. the chíldren's vóices	b. the vóices of the chíldren
	W S W S W	W S W W W S W

In (32), we would expect speakers to avoid lapse in the *of*-genitive construction of (32b) and instead prefer the eurhythmic *s*-genitive construction. Conversely, the *of*-genitive construction of the possessor and possessum pair *government* and *response* in (33b) is more eurhythmic:

(33)	a. the góvernment's respónse	b. the respónse of the góvernment
	W S W W W S	W W S W W S W W

In *the government's response* (33a), there is a lapse of three unstressed W syllables between the S syllables whereas in *the response of the government* (33b), there is only a lapse of two unstressed W syllables. Thus, we would expect that the latter construction is preferred.

Speakers, when producing one of two alternative genitive constructions, must evaluate both the *s*- and *of*- genitive forms for any given possessor-possessum pair. To model this intuition, we developed a measure of *Eurhythmy Distance* that quantifies how rhythmic each genitive construction is across the local possessor-possessum boundary. Take again the possessor and possessum pair *children* and *voices*. In the *s*-genitive construction (34a), there is one unstressed syllable between the two stressed syllables at the right and left edges of the possessor-possessum boundary:

(34) a. the chíldren's vóices

$\mathrm{W}\ \underbrace{\mathrm{S}\ \mathrm{W}\ \mathrm{S}}_{1}\ \mathrm{W}$

b. the vóices of the chíldren

$\mathrm{W}\ \underbrace{\mathrm{S}\ \mathrm{W}\ \mathrm{W}\ \mathrm{W}\ \mathrm{S}}_{3}\ \mathrm{W}$

In the *of*-genitive construction for the same possessor-possessum pair (34b), there are three unstressed syllables spanning the possessum-possessor border;[34] that is, there are three weak syllables between the leftmost stress of the possessum and the rightmost stress of the possessor.

Eurhythmy Distance takes the number of intervening syllables between stress peaks in the genitive constructions as shown in (34) above and measures how far away from perfectly alternating rhythm a given construction is. We will refer to eurhythmy distance when the possessor and possessum are in the *s*-genitive form as *s*-Eurhythmy Distance (*s*-ED) and when the possessor and possessum are in the *of*-genitive form as *of*-Eurhythmy Distance (*of*-ED). Eurhythmy Distance is calculated by taking the absolute value of the number of unstressed syllables between stress peaks across the genitive border, as formulated in (35).

[34] Assuming that the *of* is unstressed. See Chapter 6 for discussion of the potential of stressing *of* in this configuration.

(35) *s*- ED = | # of unstressed syllables between rightmost possessor stress and leftmost possessum stress – 1 |

of-ED = | # of unstressed syllables between rightmost possessum stress and leftmost possessor stress – 1 |

For the possessor-possessum pair of *children* and *voices*, then, the *s*-ED is 0, and the *of*-ED is 2.

In the Eurhythmy Distance measure, a count of 0 means that the construction exhibits the ideal eurhythmic alternation of S and W syllables, with exactly one W syllable intervening between two S syllables. Thus, any eurhythmy distance that does not equal 0 means that perfectly alternating rhythm is not achieved by the construction, and under our hypothesis, the speaker will not prefer these more arrhythmic constructions (*s*/*of*-ED > 0). Additionally, the eurhythmy distance measure makes no distinction between clashes and lapses. Compare, for example, the constructions in (36).

(36)	a. the kíd's vóice	b. the géneral's vóice
	W $\underbrace{\text{S S}}_{0}$	W $\underbrace{\text{S WW S}}_{2}$
	s-ED = | 0 – 1 | = 1	*of*-ED = | 2 – 1 | = 1

The examples in (36) have different numbers of unstressed syllables between their possessors and possessums. Despite this difference, both constructions in (36) are the same distance away from perfect rhythmic alternation (*s*-ED = *of*-ED = 1), which the eurhythmy distance measure captures. Further discussion of rhythmic clashes and lapses occurs in §5.

3.2 Control predictors

FINAL SIBILANCY. Speakers tend to avoid immediately adjacent sibilants, including [s], [z], [ʃ], [tʃ], [ʒ], and [dʒ], in an OCP-type ban on neighboring sibilant sounds (Menn & MacWhinney 1984; Zwicky 1987; a.o.). In the *s*-genitive construction, the -*'s* possessive morpheme will sometimes occur next to a final sibilant in the possessor: *the veterans + -'s + descendents*. Even though repairs such as haplology of the possessive morpheme or [ə] epenthesis exist, speakers tend to avoid the occurrence of sibilants altogether by using the *of*-genitive construction. Hinrichs and Szmrecsányi (2007) find that the presence of a final sibilant on the possessor NP significantly reduces the likelihood of the *s*-genitive in both speech and writing. After manually and automatically[35] coding for the presence of a final sibilant in the possessor NP, we found that there are significantly fewer *s*-genitives with final sibilants in their possessors (34/460) than there are *of*-genitives with final sibilants (133/663) ($\chi^2 = 34.432$, $p < 0.0001$).

ANIMACY. The animacy of the possessor is the most important single predictor of genitive construction choice in English. *S*-genitives overwhelmingly have animate possessors while *of*-genitives have inanimate ones, which has been found to be true across all studies on genitive construction choice (see especially Rosenbach 2005; 2008; Hinrichs & Szmrecsányi 2007; Szmrecsányi & Hinrichs 2008; Tagliamonte & Jarmasz 2008; a.o.). For animacy coding, we used the version of the Treebank Switchboard corpus that was annotated for animacy in the Paraphrase Link project (Bresnan et al. 2002). In this version of the corpus, almost all argument noun phrases are annotated for eleven levels of animacy (Zaenen et al. 2004) using a scheme derived from Garretson et al. 2004. We simplified these eleven levels to a binary distinction between animate entities—animals and humans—and all others, including organizations.

There are significantly more animate *s*-genitive possessors (389/460) than there are *of*-genitive ones (78/663) ($\chi^2 = 592.515$, $p < 0.0001$). *Of*-genitive possessors are more often inanimate than their *s*-genitive counterparts. The effect of animacy is strong and nearly categorical in our data; hence, the model presented in §4 includes interactions between animacy and other conditioning factors—most notably, rhythm.

SEMANTIC RELATION. As many have noted, the English genitive construction encodes a host different relations between the possessor and possessum (e.g., Taylor

[35] Automated coding was done using Python scripts and the phonological segment annotations in CMU.

1996:339–348). Following Rosenbach (2002:120–123), we collapsed several relations into a single category of 'PROTOTYPICAL' genitives, which favor the *s*-genitive, and all others into a category marked simply as 'NON-PROTOTYPICAL'. Prototypical genitives were any examples that fell into one four subclasses: kinship (*the children of these people*), body-part (*the fish's mouth*), part-whole (*the car's starter*), and physical/legal ownership (*Scotty's bed*). Tokens not fitting one of these four types were classified as non-prototypical, e.g. *an employer's rights, the owner of the store, the bag's contents*. In our data, genitives denoting prototypical relations occur as *s*-genitives (141/460) significantly more often than *of*-genitives (31/663) (χ^2 = 129.92, p < 0.0001).

THEMATICITY. Osselton (1988) examined the tendency of topical or "thematic" possessors to favor the *s*-genitive even when they are otherwise disfavored. For example, in a textbook on phonology, *sound,* which, as an inanimate possessor, would likely occur in an *of*-genitive elsewhere, would be more likely to occur in the *s*-genitive: e.g., *the sound's feature structure.* Hinrichs and Szmrecsányi (2007) found that Osselton's hypothesis holds true in written English genitives, with thematic possessors occurring more often in the *s*-genitive alternative. Following Hinrichs and Szmrecsányi (2007), we took the log text frequency of the head noun in each possessor, extracted and calculated automatically via a Python script, as a count of thematicity. We do not find a significant effect of thematicity in predicting genitive construction choice (W = 114680, p = 0.992); therefore, thematicity has been excluded from our final modeling.

GIVENNESS. It has been suggested by some that the information status of the possessor influences genitive construction choice (Biber et al. 1999; Quirk et al. 1985). When a possessor NP refers to a discourse-old entity, it is thought that a speaker is more likely to produce an *s*-genitive construction so as to place given information before new information. We manually coded givenness by looking for reference of any kind to the possessor in the preceding ten line context of each genitive token. In our data, there is a significantly greater proportion of given possessors in *of*-genitives (23%) than in *s*-genitives (12%) (χ^2 = 12, p < 0.001) in our data, counter to expectations.

WEIGHT. There is a well-known tendency in English for speakers to place "heavier" (i.e., longer and more complex) constituents after "light" (i.e., shorter) ones (Behaghel 1909; Quirk et al. 1985; Hawkins 1994; Wasow 2002; Bresnan et al. 2007; a.o.). Following much work on genitives and other English constructions (e.g., Rosenbach 2005; Bresnan et al. 2007; Hinrichs & Szmrecsányi 2007; cf. Anttila et al. 2010), we measured the weight of each possessor and possessum NP by the number of orthographic words, which has been found to so highly correlate with theoretical measures such as syntactic node count that there is little advantage to using the latter (Wasow 2002; Szmrecsányi 2004). We predict that the heavier the possessor NP is, the more likely it will follow the possessum and occur in the *of*-genitive form, as in the extreme example in (37).

(37) a. $^{\checkmark}$[the attitude]$_{\text{possessum}}$ of [people who are really into classical music and feel that if it's not seventy five years old, it hasn't stood the test of time]$_{\text{possessor}}$

b. $^{???}$[people who are really into classical music and feel that if it's not seventy five years old, it hasn't stood the test of time]$_{\text{possessor}}$'s [attitude]$_{\text{possessum}}$

The *of*-genitive in (37a)—a token from our data—exemplifies the end weight effect, with an unusually long possessor (24 words) following a short, one-word possessum. In (37b), which is a construct, the longer possessor precedes the one-word possessum, making this alternative dispreferred.

It is important to note here that our measure of end weight is neither a strictly syntactic nor a strictly phonological measure. Some researchers have framed end weight as a phonological property, using, for instance, the number of syllables for constituent weight (e.g., McDonald et al. 1993). the number of prosodic phrases (Zec & Inkelas 1990), or the number of lexical stresses as a measure of phonological and prosodic complexity (e.g., Anttila et al. 2010). Others have assumed weight effects to be driven by the demand of processing complex syntactic structures and measure weight in terms of the number of syntactic nodes or dependencies (e.g., Hawkins 1994; Gibson 2000; Temperley 2007). In a study comparing different phonological, processing, and syntactic measures of weight, Grafmiller & Shih (2011) present evidence suggesting that weight measured in orthographic words is a less ideal proxy for either phonological or syntactic notions of weight, especially in the English genitive alternation. They note, however, that absolute differences in the explanatory power of the different weight measures investigated are small, and the ranking of the weight factor relative to other factors in their models did not change regardless of which weight measure was implemented. They conclude that any differences in the predictive power of word count versus other measures are likely to be of little import to studies focusing mainly on other factors, including the present one. Since orthographic word count is by far the simplest measure to operationalize, we follow the precedence in construction choice studies by using word count in this study instead of other measures of end weight (see also Szmrecsányi 2004).

As expected, we find that the mean number of words in the possessor is significantly lower among *s*-genitives (*mean* = 1.8, *SD* = 0.62) than among *of*-genitives (*mean* = 2.62, *SD* = 2.31)(W = 114680, $p < 0.0001$). To operationalize weight for the study presented here, we use the difference in word count between the possessum and possessor phrases (i.e., log(possessor word count) – log(possessum word count)) (Grafmiller & Shih 2011), following similar methodology from Bresnan & Ford (2010). The difference provides a scale of weight measurement, whereby the more positive the difference, the longer the possessor phrase is, and the more negative the difference, the longer the possessum phrase is. We expect that longer possessors (more positive weight measures) will prefer the *of*-genitive construction, and longer possessums (more negative weight measures) will prefer the *s*-genitive construction.

PERSISTENCE. Persistence describes a possible priming effect of one structure on subsequent construction choices. For example, in genitive construction choice, the presence of an *s*-genitive may prime the choice of another *s*-genitive the next time the speaker has to choose between constructions. Previous genitive research (Szmrecsányi 2006; Hinrichs & Szmrecsányi 2007) has found persistence to be a significant but small effect in both spoken and written English. While we excluded pronominal genitives (see §2), which meant that we could only easily calculate persistence based on genitives without pronouns, we nevertheless found a significant difference in the proportions of *s*-genitives (32.2%) and *of*-genitives (18.2%) that are immediately preceded by another *s*-genitive in our data ($\chi^2 = 28.35$, $p < 0.0001$).

SPEAKER AGE AND GENDER. Since around the 16th century, the frequency of the *s*-genitive has been steadily increasing (Rosenbach 2007:154), and this trend has continued through the latter half of the 20th century in both American and British English (Hinrichs & Szmrecsányi 2007). Because of its French origins and predominance prior to the 16th century, the *of*-genitive form is often regarded as having formal connotations (Rosenbach 2002; Tagliamonte & Jarmasz 2008). This has led some to hypothesize that women, who have been found in sociolinguistic studies to utilize formal structures more frequently than men, along with people with higher education in general, are more likely to use *of*-genitive constructions. In their study on spoken genitives in Toronto English, Tagliamonte & Jarmasz (2008) show a correlation between older age and the use of more *of*-genitives but do not find significant effects of speaker gender or education. We utilize the speaker sex and age information available with the Switchboard data. Speaker ages ranged from 19 to 67, with a median age of 35. Speaker education was excluded due to missing educational information for some of the subjects.

4 Modeling and analysis

In this section, we present a model of genitive construction choice in spoken English using a logistic regression analysis and the conditioning factors presented above.[36] In addition to the final model presented below, three mixed-effects models containing combinations of speaker and conversation as group levels (random effects) were also

[36] Graphics and statistics were prepared using the R statistical computing platform (R Development Core Team 2010) and the Design library (Harrell 2009).

tested: speaker only (N = 922), conversation only (N = 461), and both speaker and conversation. In none of the mixed effects models did any of the grouping factors account for a significant portion of the variance in our data. This is likely due to the fact that the large majority of individual speakers are represented by only one or two genitive tokens in our dataset (52% and 28%, respectively). Individual conversations are similarly underrepresented, with 80% contributing three or fewer tokens. We therefore considered the use of a simpler single-level model for our final analysis to be justified, following other recent work (e.g., Hinrichs & Szmrecsányi 2007; Tagliamonte & Jarmasz 2008).

Factors in the final model were selected via stepwise backward elimination in which insignificant factors were removed sequentially from the full model containing all of the previously described factors in §3. The criterion for removal of predictors was if and only if the absolute value of the coefficient was less than twice the standard error. For all models, binary predictors were centered by subtracting the mean and numerical predictors were centered and standardized by dividing by twice the standard deviation. Centering and standardizing predictors protects against harmful effects of data multicollinearity, and normalizing numerical predictors by two standard deviations allows us to directly compare their model coefficients with those of binary predictors (Gelman 2008).

Table 3. Logistic regression estimates: Ratios represent the relative chances of *s*-genitive over *of*-genitive

Factor	Estimate	Std. Error	Z value	Pr (>|z|)	
Intercept	-0.6863	0.1031	-6.66	<0.0001	***
Possessor animacy = inanimate	-3.7161	0.2116	-17.56	<0.0001	***
Word count (log diff)	-3.3216	0.5578	-5.96	<0.0001	***
Final sibilant	-1.1525	0.3075	-3.75	0.0002	**
Semantic relation = prototypical	1.042	0.3044	3.42	0.0006	**
s-ED_{ph}	-0.1823	0.658	-0.28	0.7818	
of-ED_{ph}	0.066	0.2333	0.28	0.7774	
Possessor givenness	0.4483	0.259	1.73	0.0835	

= not given					
Possessor freq (log)	-0.0262	0.233	-0.11	0.9104	
Persistence	0.3568	0.2186	1.63	0.1025	
Speaker birthdate	0.0037	0.0018	2.08	0.0375	.
Speaker sex = M	-0.3433	0.1933	-1.78	0.0757	
Interactions					
s-ED_{ph} * animacy = inan	2.6076	1.2856	2.03	0.0425	.
of-ED_{ph} * animacy = inanim	1.9599	0.4663	4.2	<0.0001	***

N	1107	*of* (653) / *s* (454)	
model χ^2	748.81	R^2	0.663
Dxy	0.839	%correct (%baseline)	92 (69.53)
κ	1.797053	AIC_c	778.2298

. significant at $p < 0.05$, * significant at $p < 0.01$, ** significant at $p < 0.001$, *** significant at $p < 0.0001$

Our model accurately predicts 92% of the data and accounts for about two-thirds of the variance in the dependent variable (R^2 = 0. 663). The model exhibits low multicollinearity (κ = 1.797053), indicating that there is no harmful overlap amongst multiple predictors with respect to the variance that each explains. In general, κ values below 6 suggest little to no multicollinearity (Baayen 2008:182).

The models were verified using a step-up method where each predictor, beginning with those previously identified as significant in the literature, was added one at a time until no further improvement of the models occurred. We then tested the model for over-fitting using bootstrap resampling (*N* runs = 1000) in which the model was fit to random resamples of the dataset. Table 3 provides the results of the model.

Table 3 above reports only the size and direction of the predictor effects in the model. Figure 7 shows the explanatory power that each predictor has in the model, measured by the difference in -2 log likelihoods between two nested models. To calculate each predictor's explanatory power, we removed each predictor from the full model. The decrease in the model's goodness-of-fit (increase in -2 log likelihoods) was recorded with each predictor removed in turn.

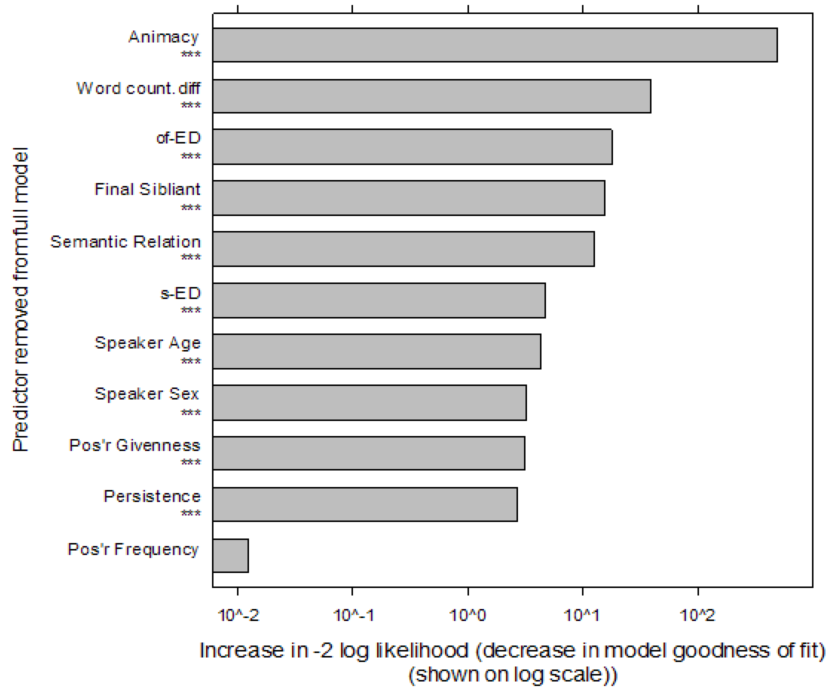

Figure 7. Increase in -2 log likelihood (decrease in model goodness-of-fit) if factor removed.

As is evident from Figure 7 animacy holds the most explanatory power for our data. The explanatory power of the next most important predictor—possessor weight—follows far behind the power of animacy in predicting genitive construction choice. Eurhythmy distance in the *of*-genitive form, final sibilants in the possessor, and semantic relation each independently make significant, though minor, contributions to the model, while possessor head frequency makes no contribution to the model fit.

This finding parallels the results of other recent work on genitive construction choice (Hinrichs & Szmrecsányi 2007; Szmrecsányi & Hinrichs 2008; Tagliamonte & Jarmasz 2008) that found possessor animacy, possessor length, and final sibilants on the possessor NPs to be reliable predictors in their models . We find that semantic relation, the age of the speaker (by birthdate), and the givenness of the possessor also reliably predict genitive choice in our data. Figure 8 shows the partial effects plots of all our predictors, with the exception of rhythmic interactions with animacy, which are discussed separately below. In each graph, a greater log odds value (*y*-axis) indicates an increased probability of an *s*-genitive for the given value of that predictor (*x*-axis) when all other predictors in the model are held constant.

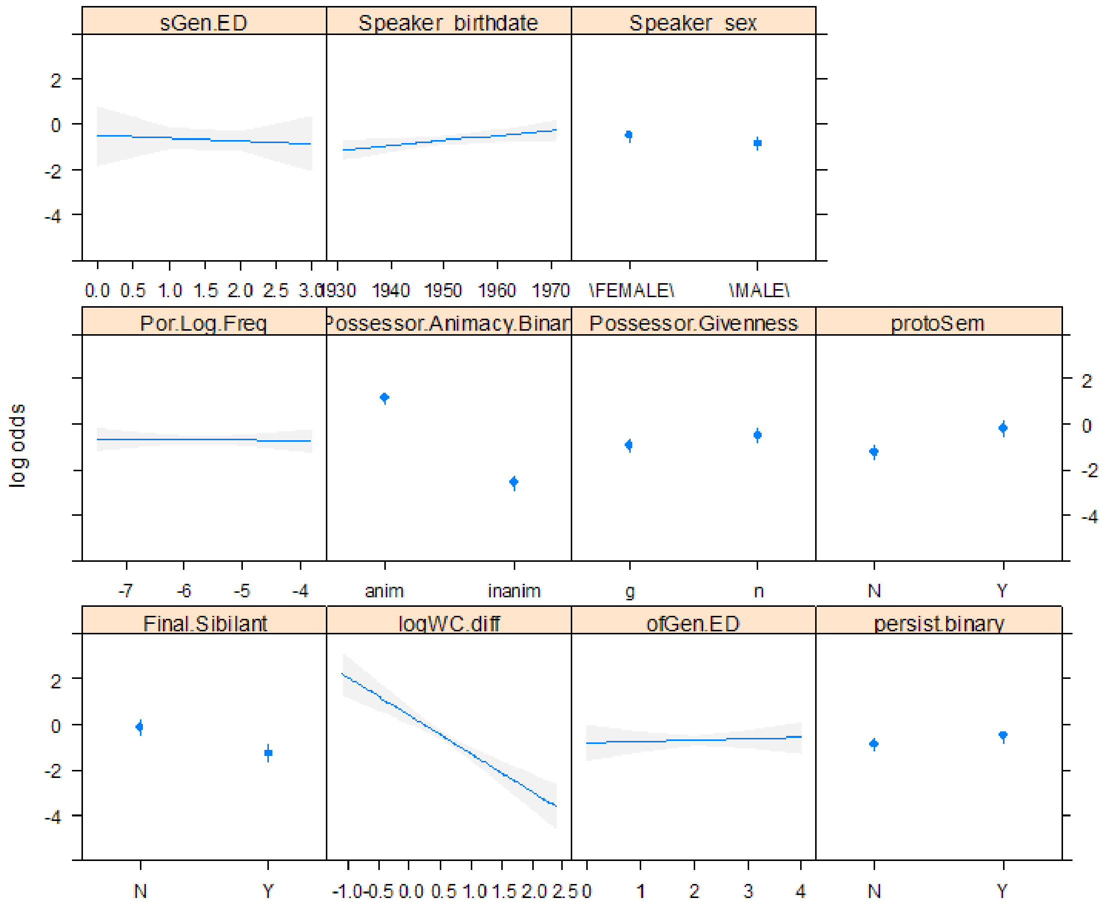

Figure 8. Partial effects of model predictors

All else being equal, we find that inanimate possessors are 2% as likely as animate possessors to occur in the *s*-genitive (see Figure 8). Also in our data, we find a roughly 40-percent decrease (odds ratio = 0.414) in the likelihood of an *s*-genitive as the possessor word count increases from one word to six or more (two standard deviations of word length). This effect of possessor length is reflected in Figure 8. Figure 8 demonstrates that the OCP avoidance of adjacent sibilants in the possessor significantly affects syntactic word order. When a possessor ends with a sibilant, the *s*-genitive alternative is only about 30-percent (odds ratio = 0.304) as likely as the *of*-genitive. In accord with our analysis of the relative frequencies of given possessors across the two constructions, we find that there is a weak but significant tendency for discourse-given possessors to favor the *of*-genitive construction. Semantic relation also plays a role: prototypical genitives are over three times as likely to occur in the *s*-genitive. Finally, speaker age behaves as hypothesized: younger subjects tend to use the *s*-genitive form more than older subjects, as is evident the increase in log odds as the birth date becomes more recent.

Finally, we come to the effects of eurhythmy distance. Figure 9 provides the partial effects plots of the interaction of eurhythmy distance with possessor animacy, with all other predictors in the model held constant. *Of*-ED exhibits a significant interaction with animacy, indicating that the animacy of the possessor has a significant effect on the influence of rhythmicity in determining genitive choice. The interaction of *s*-ED and animacy is not significant, but it did not meet our criterion for removal from the model. Both are discussed in turn.

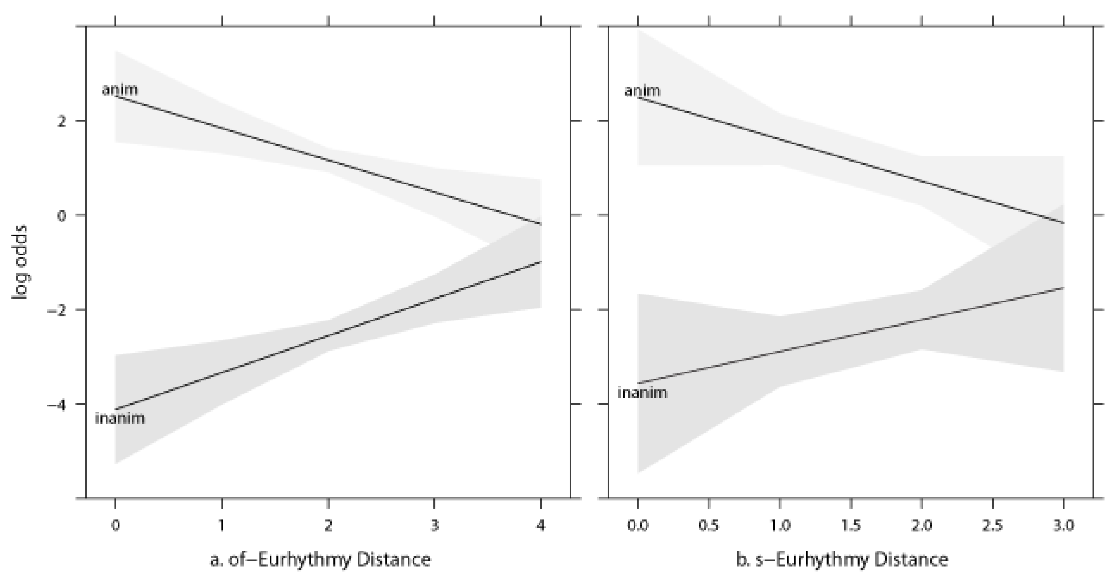

Figure 9. Log odds of ED measures by Possessor Animacy.

For *of*-ED, we predicted a positive slope based on the hypothesis that as *of*-ED increases—that is, the further away from eurhythmy the *of*-genitive gets—the more likely an *s*-genitive should occur to avoid rhythmic violations. The cumulative effect of *of*-ED and its interaction with animacy has a positive slope in genitives with inanimate possessors (0.2333 + 0.4663 = 0.6996), as seen in Figure 9a. However, we find that in genitives with animate possessors, *of*-ED does not have a reliable predictive value. This is reflected graphically in Figure 9a, in which the confidence interval of the predicted odds in *of*-ED with animate possessors crosses 0, thus indicating an even chance for the choice of either construction. *Of*-ED, therefore, is a reliable predictor of genitive construction choice only when the possessor is inanimate.

The interaction between *s*-ED and animacy is marginally significant, and the effect is in the expected direction. Amongst animate possessors, as the distance from perfectly alternating rhythm grows in the *s*-genitive construction, there is a trend away from the *s*-genitive; however, we find that the confidence interval crosses 0, indicating that the model does not reliably predict an outcome for this factor. As with *of*-ED and animate possessors, there is a slightly positive slope of *s*-ED amongst inanimate possessors, but this upward trend towards the *s*-genitive is not significant, which is graphically evident from the wide and over-lapping confidence intervals.

Animacy is clearly such a strong predictor of the genitive alternation that it heavily influences the effect of rhythmicity on construction choice. We should note that, given more data, we might and expect to see the emergence of a stronger effect of *s*-ED as well as animacy-independent effects of rhythm. Other research has also shown close ties between semantic factors and rhythmic alternation (e.g., Hanssen et al. 2013), suggesting that the interaction between these two domains is non-trivial. Furthermore, the interaction between rhythm and animacy factors here contributes to the explanation of the relative rarity and weakness of phonological conditions on syntactic processes: because phonological factors must compete with higher-order factors, we naturally expect their effects to be small. That is, given greater weighting of importance for higher-order factors, the cost is lower if a phonological factor such as rhythm has to be violated in order to satisfy, for example, semantic preferences.

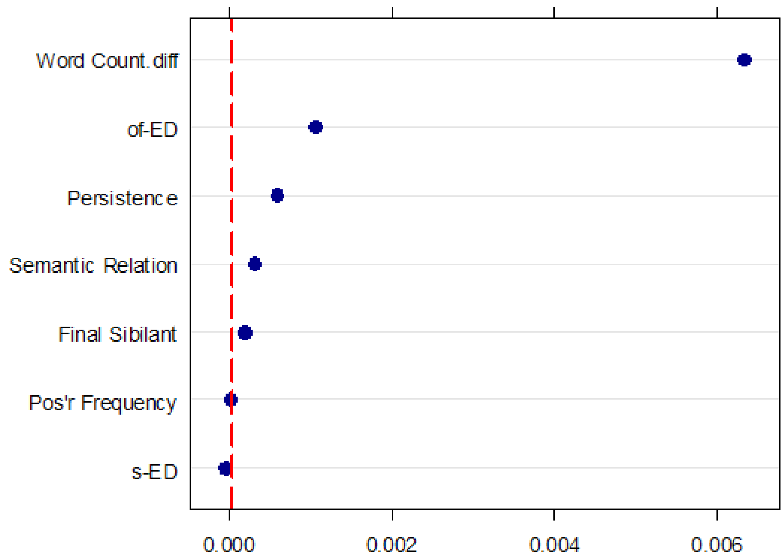

Figure 10. Variable importance in genitive choice (animacy not shown). Predictors to the right of dashed vertical line are significant.

4.1 Variable importance via conditional random forest analysis

Our results are verified using conditional random forest analysis. Random forest analysis is a non-parametric, classification tree-based classifier. It allows for robust testing of individual predictor contributions via randomized and conditional permutation tests over forests of classification trees with subsets of both the test data and the test variables (Breiman 2001; for conditional random forests, see Strobl et al. 2008; Strobl, Hothorn, et al. 2009; Strobl, Malley, et al. 2009; Hothorn et al. 2013; a.o.). More description of conditional random forest methods and applications within linguistics are found in Grafmiller & Shih (2011); Tagliamonte & Baayen (2012); a.o. Analysis was performed using the `R` package `party()` (Hothorn et al. 2013).

Variable importance measures from the results of a conditional random forest analysis on the genitive model is given in Figure 10.[37] A full table of results is provid-

[37] Animacy is the strongest predictor (varimp = 0.176731), and is not shown here.

ed in Appendix A. The results of the random forest corroborates the results of the logistic regression model presented in §4, and demonstrates *of*-ED makes an important individual contribution to predicting genitive construction variation.

5 Discussion

In the previous section, we presented our model of genitive construction choice in English using *s*- and *of*-eurhythmy distance to quantify rhythm (henceforth Model I). In this section, we discuss the efficacy of our eurhythmy distance measure in comparison to more standard and separate rhythmicity measures of clash and lapse (§5.1). Then we consider the differences between *s*-ED and *of*-ED (§5.2). In §5.3, we present a study of a *Comparative Eurhythmy Distance* measure that combines both *s*-ED and *of*-ED into a single predictor and argue for the necessity of both eurhythmy distance measures.

5.1 Eurhythmy Distance versus Clash and Lapse

One departure of our eurhythmy distance measure presented here from previous treatments of rhythmicity (e.g., Anttila et al. 2010) is the collapsing of the clash versus lapse distinction in the ED count. ED does not differentiate between clash and lapse: both are considered one step away from perfectly alternating rhythm (*s*/*of*-ED = 1). In the previous literature, however, it is a common hypothesis that stress clash is more grave a violation of the Principle of Rhythmic Alternation than stress lapse. Nespor and Vogel (1989:87) state: "while there is a strong tendency to eliminate lapses, they are not felt to be quite as disturbing as clashes." This suggests that the loss of distinguishing clash and lapse should be a costly one.

We can examine the actual effect of clash and lapse by substituting these measures for ED in an otherwise identical model of genitive construction choice (henceforth Model II). Model II includes clash in the *s*-genitive form (no unstressed syllables intervening between stressed syllables at the possessor-possessum border); *s*-genitive lapse (the distance away from perfectly alternating rhythm if there are two or more unstressed syllables between stress peaks at the possessor-possessum border); and *of*-genitive lapse (the distance away from perfectly alternating rhythm if there are two or more unstressed syllables between stress peaks at the possessum-possessor border). Clash in the *of*-genitive is unnecessary because *of* is treated as unstressed and as such, *of*-genitives will never have stress clash. The model also includes the interactions of these rhythmic predictors with animacy, as in Model I.

Holding non-rhythmic predictors constant, we find that neither *s*-genitive lapse (β = -0.102, z = 0.39, p = 0.699), *of*-genitive lapse (β = -0.310, z = 1.19, p = 0.24), nor *s*-genitive clash (β = 0.218, z = 1.19, p = 0.35) are good predictors of genitive construction choice on their own. The interactions between animacy and lapse for both *s*- and *of*-genitives were significant (for *s*-genitive lapse: β = 1.56, z = 3.01, p = 0.003; for *of*-genitive lapse: β = 1.542, z = 3.10, p = 0.002), but the interaction between animacy and *s*-genitive clash is not significant and falls within our criterion for removal from the model (β = 0.249, SE = 0.466, z = 0.53, p = 0.593).

This result runs counter to the hypothesis that stress clashes are more disfavored than lapses and that, in the event of stress clash, the alternative construction will be chosen. A possible explanation for the unreliability of stress clash as a predictor is that speakers have other repairs available to avoid clash: the Rhythm Rule, for example, in environments where stress shift or retraction may occur.[38] Since our rhythm counts only include dictionary-based lexical stress, there is no way to know for sure without consulting the actual Switchboard sound recordings, which is left for future study. What speakers may be doing in the presence of clash is repairing the clash via stress shift, retraction, or promotion. Lapse, unlike clash, may be more difficult to correct since stress insertion on unstressed syllables is a less likely repair (though cf. Selkirk

[38] While the Rhythm Rule does not make a stressed syllable into a completely unstressed one, it may be possible that we would find Rhythm Rule-like phonetic enhancements or reductions (see e.g. Chapter 6).

1984). Hence, we see from Model II a clear influence of stress lapse, where longer lapses in stress will result in speakers choosing the alternative construction.

The unreliability of stress clash might also provide an explanation for the low performance of the *s*-ED measure in Model I. The *s*-ED measure incorporates clash as well as lapse. In Model II, we see that clash holds no explanatory power for genitive construction choice; thus, the incorporation of clash in the *s*-ED measure might weaken its effect. The combined ED measure with both clash and lapse, on the other hand, allows us to capture the influence of rhythmicity with fewer degrees of freedom than clash and lapse, thereby preventing over-fitting of the model from too many predictors and potentially high multicollinearity amongst factors.

5.2 *s*-ED versus *of*-ED: prosodic phrasing

The approach to quantifying simple alternating stressed and unstressed syllables utilized in this paper departs from much of the previous literature on rhythm and syntax interaction, which focuses on phrasal and prosodic stress and phonology. Given that the *s*- and *of*-genitives have different prosodic and syntactic structures, we might expect to see these differences reflected in rhythmicity's influence on genitive construction choice—particularly in how strictly the Principle of Rhythmic Alternation applies within different prosodic domains (see esp. Selkirk 1984; Nespor & Vogel 1986; a.o.). Within a single prosodic phrase, language users have been noted to desire greater eurhythmy than across prosodic phrase boundaries. For example, certain stress shifting repairs such as the Rhythm Rule in English operate only within noun phrases and not without. To illustrate, consider the sequence of *thirteen* and *men* in (38).

(38) a. In the room, there were *thirteen men*.
b. When he was *thirteen*, *men* seemed much smarter to him.

In (38a), *thirteen men* is one prosodic phrase; therefore, the Rhythm Rule applies to avoid the stress clash of *thirTEEN* and *MEN*, and the main stress of *thirteen* shifts to the first syllable, forming perfectly alternating stress: *THIRteen MEN*. In (38b), the sequence *thirteen men* does not form a single prosodic phrase, and the stress clash is not repaired via the Rhythm Rule.

The genitive constructions exhibit a difference in prosodic domains (39). The *s*-genitive construction forms a single NP and prosodic phrase.

(39) a. [the car's wheel]$_{\phi}$ b. [the wheel]$_{\phi}$ [of the car]$_{\phi}$

On the other hand, two prosodic phrases form the *of*-genitive (10b). The prosodic phrasing in (39) is independently corroborated by the presence of speaker disfluencies in the Switchboard genitive dataset. We hand-coded for disfluencies, as in (40), intervening between the possessor and possessum of genitive constructions in our spoken data.

(40) a. the norms of, *um*, public behavior
b. the school district's, *you know*, goals

Our data indicates that speakers insert significantly more disfluencies in *of*-genitive constructions (n=79) than in *s*-genitive constructions (n=24)(χ^2 = 12.316, p < 0.001). The greater number of disfluencies in the *of*-genitive can be taken as evidence for the looser prosodic and phrasal constituency in the *of*-genitive constructions. Conversely, speakers insert fewer disfluencies in *s*-genitives because they have tighter prosodic constituencies.

Because of the difference in prosodic phrasing between the *s*- and *of*-genitives, a phrase-oriented approach would predict that the Principle of Rhythmic Alternation applies more strictly within *s*-genitives, which are singular prosodic units, and for eurhythmy distance in *s*-genitives (*s*-ED) to be the most—and perhaps only—important factor when speakers consider alternative constructions. Our model, however, demonstrates the opposite result: while *s*-ED is not a reliable predictor of construction choice—it only trends in the correct direction—*of*-ED is a reliable predictor, suggesting that, despite a difference in the prosodic phrasing of the genitives, the difference is not reflected in the effect of rhythmicity on construction choice. Irrespective of higher level stress domains, our results show that even the low-level and simple binary alternation of stressed and unstressed syllables influences speaker choice of syntactic ordering.

5.3 Eurhythmy Distance versus Comparative Eurhythmy Distance

In addition to the measure of eurhythmy distance, we also developed a measure of *comparative eurhythmy distance* (CED), which incorporates a comparison between *s*-ED and *of*-ED. Because our hypothesis is that speakers choose the alternative construction when the ED of either the *s*- or *of*-genitive is not equal to 0, we wanted to compare how rhythmically optimal one genitive construction is over the other, mimicking, in a sense, the same (unconscious) decision process that a speaker might undergo when making a construction choice.

To calculate CED, we use the formula CED = *of*-ED – *s*-ED. The resulting measure provides a scale wherein the more positive the CED, the more eurhythmic the *s*-genitive alternative is, and the more negative the CED, the more eurhythmic the *of*-genitive is. The comparative eurhythmic distance measure is an attractive one because, unlike the simpler measure of eurhythmy distance, CED reflects a weighing of the two potential genitive constructions against each other to predict which one the speaker is more likely to choose. This type of approach is not dissimilar to the scale that Kendall et al. (2011) and Bresnan & Ford (2010:174) develop for comparing the syntactic complexity of themes and recipients in studies of dative construction choice (see also Grafmiller & Shih 2011 and weight measure described here in §3.2).

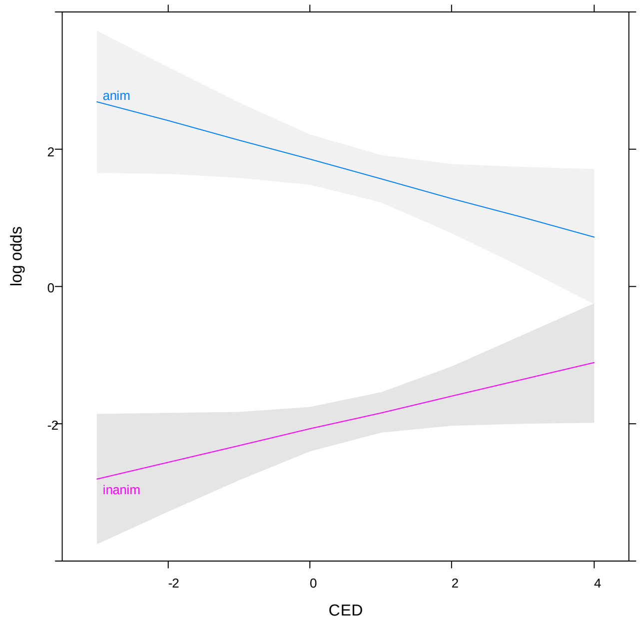

Figure 11. Log odds of Comparative Eurhythmy Distance by animacy of possessor, holding all other predictors constant. (Higher log odds value indicates greater likelihood of *s*-genitive; lower log odds value indicates greater likelihood of *of*-genitive.)

We investigated a model using Comparative Eurhythmy Distance as a measure of rhythmic influence on genitive construction choice in lieu of eurhythmy distance (henceforth, Model III). All other predictors in the model remained the same as in §4. The results of Model III are similar to those in the model presented in §4. Like ED, CED exhibits a significant interaction with animacy (β = 1.109, z = 2.84, p = 0.005). See partial effects plot in Figure 11 The cumulative effect of CED and its interaction with animacy produces a positive estimate slope in genitives with inanimate possessors (0.102 + 1.109 = 1.211), indicating that amongst inanimate possessors, speakers are more likely to choose *s*-genitive forms as CED increases and *of*-genitive forms as CED decreases. Amongst animate possessors, however, the influence of CED on genitive construction choice is unreliable. This result differs from the animacy findings in a model with separate *s*- and *of*-ED measures.

In essence, the Eurhythmy Distance and Comparative Eurhythmy Distance measures are similar, both based on counting the number of rhythmic violations a genitive construction incurs—that is, how far from perfect rhythmic alternation a given genitive is. Diverging from the simpler eurhythmy distance measure, comparative eurhythmy distance is a relative quantification intended to characterize the choosing of a more rhythmically optimal genitive by speakers. CED collapses the two ED measures, but in doing so, it obscures the disparity in how animacy interacts with *s*- and *of*-ED. The distinction between *s*- and *of*-ED's predictive value amongst animate and inanimate possessors, respectively, is important because it demonstrates that, in spoken English genitive construction choice at least, low-level rhythmic effects are subservient to stronger semantic predictors like animacy. One might imagine that speakers are predisposed to either the *s*- or *of*-genitive form based on the animacy of the possessor: animate possessors strongly prefer the *s*-genitive construction while inanimate possessors prefer the *of* construction. Rhythmic costs are weighed in terms of these animacy preferences: within genitives with animate possessors, the *s*-genitive is evaluated for optimal rhythm because animate possessors favor the *s*-genitive, and within genitives with inanimate possessors, the *of*-genitive is evaluated for optimal rhythm because inanimate possessors favor the *of*-genitive form. The alternative form surfaces if either the *s*- or *of*-genitives in animate or inanimate constructions, respectively, have stress patterns that are too deviant from perfectly alternating rhythm. Animacy's interactions with eurhythmy distance suggest that the consideration of high-level (semantic) predictors like animacy constrains the consideration of lower-level factors like the Principle of Rhythmic Alternation.

The measure of comparative eurhythmy distance offers us an elegant calculation to quantify a speaker's choice between the rhythm of the *s*-genitive construction and the parallel *of*-genitive construction. The sheer amount of information that comparative eurhythmy compresses into a single count is valuable for logistic regression-based studies, as too many predictors potentially cause harmful over-fitting of the data. In its current state, however, CED fails to encode significant detail and inequalities between *s*- and *of*-genitives and their interaction with animacy in possessors. The measure of ED, in comparison to CED, is simpler and provides more granularity for witnessing the effects of animacy and rhythmicity interaction in the two different genitive constructions. Our model presented in §4 suggests that there is independence between the two *s*- and *of*-genitive measures in their interactions with animate and inanimate possessors; thus, two independent measures should be utilized.

6 Conclusion

We began this study with two major questions about the role of rhythm in genitive construction choice in spoken English: [1] How good is rhythm as a predictor of genitive construction choice?, and [2] How important are rhythmic influences when combined with other phonological and non-phonological predictors? To answer these questions, we utilized a method of quantifying rhythm developed herein and in Chapter 3: eurhythmy distance. Our regression and random forest analyses show that ED plays a significant role in genitive construction choice, but the role that it plays is small in relation to other factors and is conditioned and constrained by the effect of animacy. Such interaction with animacy provides one explanation for the relative rarity of syntactic choices as a phonological repair for non-optimal rhythmic patterns: because animacy conditions are weighted so highly in syntax, lower-weighted phonological preferences will have less of an effect in influencing extra-phonological operations.

The exploration of rhythm in spoken construction choice and the development of the measure of eurhythmy distance in this paper are amongst the first of their kind; therefore, there is great necessity for further work and refinement (see Ehret 2011; and Grafmiller forthcoming, which build on this approach for written language). In this study, we only consider a local measure of rhythm, looking with limited scope at the boundary between possessors and possessum, under the assumption that choice points hold the greatest potential for rhythmic (and psycholinguistic) optimization (Jaeger 2010; Kuperman & Bresnan 2012; a.o.). This narrow and short-sighted vision of rhythm may be too myopic. Rhythm, from Abercrombie (1967), is defined as the expectation of regularity in evenly spaced stresses, so a more accurate measure of rhythm might be a more global one with wider scope, testing whether the rhythmic regularity expected by the language use is maintained throughout the genitive construction by using one genitive over its alternative form (see also Temperley 2009 for a longer view of rhythmic regularity).

We have also utilized idealized, dictionary-based stress annotations for the purposes of this study, which were hypothesized to reflect speakers' stored lexical representations. The actual phonetic pronunciations, however, may or may not follow dictionary approximations (see Chapter 5). In the actual spoken stream, it is possible that greater or lesser effects of rhythmicity may be found, especially taking into account repairs of stress violations via the Rhythm Rule (or Rhythm Rule-like phonetic compensation) or the elision of unstressed syllables in rapid speech (Kaisse 1985). The spoken Switchboard data used in this study provides an opportunity for future phonetic verification and investigation of the current results.

There are many further avenues of research that are necessary to better understand rhythm's role in genitive construction choice. For instance, the role of rhythm in spoken and written construction choice may differ due to the natures of spontaneous speech and calculated writing. In writing, speakers have potentially more time to consider the alternatives between *s*- and *of*-genitives, which could result in a greater effect of rhythm on construction choice (see e.g., Tilsen et al. 2012 for speech planning effects on rhythmic regularity). At the same time, writers may not be as concerned with phonological properties in written work, and rhythm's effect may diminish in comparison to spoken use (Grafmiller forthcoming). Such a diminished effect is especially expected in contexts that have overriding external constraints, such as orthographic space limitations in newspaper texts. Whether the role of rhythm in genitive construction choice has changed throughout the development of English or between the world's dialects of English also may provide further understanding of what influences speakers to make the choice of one genitive construction over another (e.g., Ehret 2011).

The results of this study demonstrate that rhythm should be considered a potential influencer of construction choice in English. Its role is small, especially when compared to known semantic, pragmatic, processing, and other phonological factors. Rhythm, as a dependent on other predictors, also does not have complete explanatory power over construction choice—nor do we expect such a strong effect, given that morphosyntactic operations are necessarily driven by semantic targets and other higher-order considerations. Though its role may be small, rhythmicity still does participate in the decision between genitive construction alternatives, providing an example of metrical conditioning of "periphrastic"—to use the term broadly—morphosyntactic alternation.

Chapter 5. Prosodification of word categories

1 Introduction

Thus far in the dissertation, we have seen that rhythmic well-formedness can condition word and syntactic construction choices in English. Moreover—though not the primary focus—it has also been shown that segmental effects such as OCP and homophony avoidance affects morphosyntactic choices as well (see Chapters 3 and 4). These results have demonstrated that morphosyntactic constructions compete to optimize (in part) phonological structure.

The question that follows from the potential of interaction between phonology and morphosyntax is how far does such interaction go? What amount of information—phonetic, phonological, morphosyntactic—is available at once? One extreme view—the null hypothesis—would be that little to no architectural separation exists between components of language, so interaction between phonetic, phonological, and morphosyntactic information should exist. The current and following chapters address this question through the test case of rhythm and stress in lexical (content) versus grammatical (function) words. Because content and function words represent differing levels of phonological and phonetic coding in their stress properties (e.g., Selkirk 1984), they can be used as a natural test case with which to diagnose the amount of interaction between phonological, phonetic, and morphosyntactic information.

The distinction between content and function words is a cornerstone of many syntactic, prosodic, and psycholinguistic theories of language (e.g., Chomsky & Halle 1968; Selkirk 1984; 1996; Chomsky 1995; Cann 2000 and references therein; a.o.). Broadly speaking, content words typically consist of nouns, verbs, adjectives, and adverbs as an "open class" of words in a language, to which new forms may be easily added. Function words, on the other hand, are those that have grammatical utility and make up a "closed class" of words in a non-expanding portion of the lexicon—for example, prepositions, articles, and conjunctions. Apart from the open and closed class difference, the division between content and function words is usually made on the basis of a number of diverse criteria, ranging from phonological generalizations to syntactic and semantic factors to neurological observations (see overview in e.g., Cann 2000).

This chapter and the next focuses on the observed phonological difference between content and function words: that function words commonly reduce in context while content words rarely do, if at all (Selkirk 1984; 1996; Kaisse 1985; Cutler 1993; Inkelas & Zec 1993; Bell et al. 2009; a.o. cf. Johnson 2004a). This behavior has been primarily attributed to an assumed fundamental difference in lexical stress properties between the two word categories: in simple terms, content words have lexically-available stress—either via lexical encoding or privileged psycholinguistic access—whereas function words do not. Function words receive their stress post-lexically, leading to reduced surface realizations that depend on various factors including morphosyntactic and prosodic structure (Kiparsky 1982; Selkirk 1984; 1996; Kaisse 1985; Inkelas & Zec 1993; a.o.).

The supposition that there are crucial differences in lexical stress for content versus function words offers a valuable arena in which to test the depth of the interface of syntax and phonology, especially with respect to the Principle of Rhythmic Alternation. Access to content word stress information only in rhythmic computation with morphosyntactic, phrasal material would reveal that there is information-sharing only between underlying phonological properties and morphosyntax. Access to both content and function word stress on the surface would reveal instead that there is information sharing of both underlying phonological properties and the output of phonological and phonetic operations that determine surface stress patterns. This issue will be addressed directly in the following chapter.

Before tackling the question of depth of information sharing between components of grammar, however, it is first necessary to examine closely the details of the assumption that content and function words are distinct with respect to reduction, stress, and prosodification. Traditionally and most commonly, the division of the lexicon into grammatical and lexical categories is taken for granted to be binary division:

all words are parceled into either the content word category or the function word category (overview in Cann 2000). Counter to this standardly-assumed binary categorization, however, some work has called into question whether the stress- and reduction-based criteria for function and content word-hood points to a binary division or whether more fine-grained partitions of the lexicon are better suited to capturing the differences in how words behave for stress, prosodification, and reduction purposes. A common additional category, for example, are clitics, which are sometimes regarded as a separate class in reduction behavior from function words (e.g., Kaisse 1985; Inkelas & Zec 1993). Noting significant durational differences, Bell et al. (2003; 2009) make a division on the basis of frequency, between high-frequency function words and mid-to-low frequency function words, in addition to the traditional content/function divide. Other work on pitch accent placement from the speech recognition realm has demonstrated that divisions using groupings of parts-of-speech and lexical specifications can better model prosodic accents or reduction in words (Altenberg 1987; Hirschberg 1993; Nenkova et al. 2007; a.o.). While the proposed, non-binary divisions of word categories range from scalar (e.g., Altenberg 1987) to discrete (e.g., Hirschberg 1993; Nenkova et al. 2007), all call into question whether the crucial lexical stress distinctions are, as standardly taken, a simple contrast between content and function words, or rather, a more complex contrast amongst several word categories. The answer to this question holds consequences for considering the differential roles of word categories within rhythmic regularity, stress, and their interface with morphosyntax.

This chapter presents a study of reduction in monosyllabic words in a corpus of conversational American English. Reduction is measured two ways: segmental reduction (e.g. *has*: [hæz] ~ [həz]) and duration. By directly examining phonetic correlates of stress and reduction, the current study fills an empirical gap in the literature on function and content word stress, which has largely focused on prosodic or pitch accent cues as measures of stress. The first part of the study proposes a set of hypotheses for the categorization of words based on observations of reduction in spoken language based on a bottom-up analysis of emergent word categories in the reduction data. The second portion of the study systematically compares the hypotheses of word categorizations proposed against previous proposals in the existing literature, including the standard binary division of content and function words. The results demonstrate that the standard binary function-content word division does not fully capture the behavior of reduction in connected speech. A large categorical divide between content and function words does exist: content words have a greater potential for stress and hence exhibit less reduction than function words. But, function words themselves are a heterogeneous set: some exhibit stress potentials similar, if not identical, to canonical content words while other words are more invariably reduced, with less potential for stress. Based on reduction patterns, the results here point to a quaternary division of the lexicon as an optimal solution, consisting of two categories containing largely prototypical function and content words and two categories that contain a mixture of content and function words that demonstrate more or less reduction than can be otherwise explained without positing additional categories.

This chapter is organized as follows. Section 2 reviews the existing literature on stress and reduction in content and function words. Section 3 makes the case that differential stress properties in word categories are not only a theoretical matter of interest but also pose crucial practical problems for studies of rhythm in naturally-occurring speech that need to be settled. Section 4 introduces the data. The two measures of reduction studied here—segmental and durational reduction—are introduced in §5, which also includes a brief look at how the standard, binary content versus function partitioning of the lexicon patterns with respect to phonetic reduction in the data. A bottom-up, cluster analysis is presented in §6, which produces hypotheses of word categorizations that emerge directly from comparing similarities and dissimilarities in reduction patterns. Control predictors that are also known to influence reduction and interact with content and function word categories are discussed in §7. Section 8 tests the competing hypotheses of word categorizations, and argues that an *n*-ary distinction of word classes—where *n* is greater than two—better fits the behavior of reduction in the data while controlling for other known predictors of phonetic reduction. The optimal hypothesis from §8 and the interactions between word categories and factors known to be sensitive to content and function word classes are examined in §9. Section 10 considers the ramifications of a quaternary division of word categories for formal and psycholinguistic theories of stress and function versus content word categorization in general.

2 Background: content versus function word prosodification

The phonological basis to a lexical versus grammatical division in the lexicon rests on the basic observation that content words do not reduce and function words do—or, put less stringently, that content words are less likely to reduce than function words. I take for granted here that the surface differences between word categories reflect deeper differences in stress encoding (Selkirk 1984; 1996; Kaisse 1985; Inkelas & Zec 1993; a.o.) or lexical accessibility (Segalowitz & Lane 2000; Bell et al. 2009; a.o.). Instead, the goal at hand is to examine the empirical basis of these assumed word category differences: starting from the surface phonetic evidence, what are the emergent word categories?

The majority view on word categories has been one of binarity (e.g., Cann 2000 and references therein): words divide (relatively) neatly into content and function categories. Surface stress of function words is then determined by a variety of conditioning environments. For example, reduction usually occurs phrase-medially (41a), while reduction is typically blocked phrase-finally (41b) (example from Inkelas & Zec 1993:209):

(41)	a.	I am sad.	✓[əm]
	b.	John is sadder than I am.	*[əm]

On the other hand, a minority view has questioned whether patterns in reduction actually reflect strict binarity between content and lexical words in the lexicon, or whether the empirical data instead points to a more fine-grained division beyond simple "content versus function."

Altenberg (1987) observed that the phonetic correlates of stress in English did not necessarily point to the simple binary division that is usually taken as a primitive given. In a prosodically-annotated subset of the London-Lund Corpus based on a 48-minute monologue (4877 words) (Svartvik & Quirk 1980). Altenberg examined a measure that he calls “zero potential.” Zero potential refers to how often a word in the corpus appeared in prosodically unmarked form, carrying little to no stress or intonational features (e.g., prominences in pitch, amplitude). Altenberg’s assumption was that the greater the zero potential of a given word (or class of words), the more likely the word would be lexically unstressed. Zero potential, then, was a measure of the word’s reducibility: a low zero potential meant that the word almost always appeared in stressed form, and a high zero potential meant that the word almost always appeared in stressless form.

Examining part-of-speech categories, Altenberg found a general divide between content and function words. Content words behaved more homogeneously in their stress behavior, with less reduction (i.e., lower zero potentials) across the set of traditional content words. The crucial difference that Altenberg uncovered, however, was in the function words. Function words proved to be a more heterogeneous set in terms of stress: some function words were like content words, with little to no reduction, whereas other function words behaved more prototypically “function”-like, in having greater zero potentials (Altenberg 1987:132). For function words, Altenberg proposed a secondary classification scale, reproduced in Table 4 below. The scale ranges from classes of words that, like content words, have a low zero potential, to classes of words that demonstrate a high zero potential. The middle of the scale includes word classes—for instance, prepositions and modal auxiliaries—that occur in roughly equivalent frequency in stressed and stressless forms in the corpus.

Table 4. Altenberg's scale of word classes (1987: 133–134)

Reducibility	Part-of-speech
0–10%	adjectives, lexical verbs, *very*, common nouns, *wh*-adverbs, ordinals, quantifying pronouns, nationality adjectives, *ly*-adverbs, compound pronouns
11–20%	phrasal-verb particles, *well*, proper nouns, demonstrative pronouns, cardinals, predeterminers
21–30%	closed-class adverbs, other subordinators, *do* (all forms and uses)
31–40%	*have* (all forms and uses)
41–50%	relative pronouns, modal auxiliaries, quantifying postdeterminers, demonstrative determiners (plural)
51–60%	prepositions, *and*, correlative subordinators
61–70%	subordinator *that*, *but*, demonstrative determiners (singular)
71–80%	*be* (all forms and uses), personal pronouns
81–90%	possessive determiners, existential *there*
91–100%	infinitive marker, definite article, indefinite articles

Following Altenberg's work, Hirschberg (1993) examined several speech corpora of American English, looking at how often words carried pitch accent. Hirschberg found an improvement of pitch accent prediction from 68 to 77% accuracy when finer-grained word categories, as given in Table 5, were used, instead of a binary content versus function word grouping (321–322). The four categories that Hirschberg arrived at were generated via manual optimization.

Table 5. Hirschberg's four-way division of word classes (1993: 321)

Word class	Part-of-speech
open	nouns, verbs, adjectives, adverbs (except any specified below)
closed_accented	negative article, negative modals, negative *do*, most nominal pronouns, most nominative and all reflexive pronouns, pre- and post-qualifiers, pre-quantifiers, post-determiners, nominal adverbials, interjections, particles, most *wh*-words, some prepositions
closed_deaccented	possessive pronouns, *wh*-pronouns, coordinating and subordinating conjunctions, existential *there*, *have*, accusative pronouns and *wh*-adverbials, some prepositions, positive *do*, some adverbials, nominative and accusative *it*, nominative *they*, some nominal pronouns
closed_cliticized	definite and indefinite articles, many prepositions, some modal auxiliaries, copular verbs, *have* auxiliaries

Also examining pitch accent placement in spoken American English, Nenkova et al. (2007) find that part-of-speech categories are predictive of pitch accent. This is not directly contrasted to a binary division, but it provides further evidence that a binary division of content versus function words is not necessarily the only explanatory option of surface prominence patterns.

3 Function word classification: A methodological issue

Setting aside temporarily the theoretical issues of content versus function word prosodification, this section turns to a related practical and methodological issue facing any study of rhythmic patterning across naturalistic, connected speech: distinguishing content and function words. The categorization of content and function words leads, as discussed in §2, to crucial differences in the assumption of lexical stress. Since many standard theories take content words to have lexical stress and function words to not, then the first step must be to determine what is a content word and what is a function word.

The task of separating content and function words seems straightforward when viewed on a broad scale. Traditionally, content words are considered to be nouns, verbs, adjectives, and—under some definitions—adverbs. They have been defined as the set of "open class" items, in that new content words may be added to the language, either through derivation, borrowings, or semantic changes. Content words can also be morphologically complex. Function words, on the other hand, are typically considered to be the set of words that content words are not: that is, everything except nouns, verbs, and adjectives. Function words are a "closed class," meaning that new items may not be added to this set, and they are morphologically simplex. Furthermore, another diagnostic of content versus function word-hood has been stress: content words carry stress with no reduction, and function words are stressless and can occur in reduced forms.

In practice, however, exhaustively partitioning the lexicon into content versus function words is no trivial task. When examined closely, the typical criteria, as outlined above, for distinguishing content and function words is not entirely deterministic (see Cann 2000 and references therein for an overview). One criterion for function word-hood is an inability to undergo derivational morphology: a verb *give* can take *–er* to become *giver*, but **haver* is not an acceptable formation from the modal auxiliary *have*. But, some prepositions, like content words, can also take *–er*: *upper*, *downer*. Cann (2000:41) also notes that not all content words, as is the case with function words, undergo the expected derivational morphology: *unhappy* ~ **unsad* ~ **unmany*.

One of the most oft-touted diagnostics for function word-hood is the appeal to a distinction between open and closed classes: content words belong to the open class, and function words belong to the closed class. One issue with this criterion arises when subclasses of purportedly open class words form closed classes that do not take on new lexical items: for example, days of the week and auxiliary verbs form non-expanding, closed subclasses of nouns and verbs, respectively (Cann 2000:40). Numbers also fall into murky territory in that they are a mostly closed class, but have combinatoric productivity (e.g., seventy-four thousand) and an occasional ability to take on new items (e.g., *bajillion*, *gazillion*[39]). Furthermore, even though it is a rare occurrence, supposed closed classes of lexical items can add new members from time to time. For example, Pullum (2011) points out that certain English adjectives, such as *numerous*, *various*, and *multiple*, demonstrate recent behavior as determinatives (e.g., (42a)), even though determinatives are considered a closed class of function words (e.g., (42b) versus (42c)).

(42) a. **Numerous** of the more thoughtful teens… [from a 1996 news report]
b. **Several** of the more thoughtful teens…
c. ***Happy** of the more thoughtful teens…

A large portion of the grammaticalization literature has focused on such diachronic lexical to functional developments (e.g., Hopper & Traugott 1993).

The idea that function words form a closed class suggests that an exhaustive list of function words for any given language can be generated. Despite the common use of the content-function word divide—particularly in natural language corpus work—, one is still hard-pressed to find a consistent, comprehensive master list of grammatical words for English.[40] Typically, the classification of content and function words is undertaken using stipulated assumptions that vary from researcher to researcher. Furthermore, specific details of the division are commonly not reported, even when they are crucial to the study. For example, Johnson (2004a), which examines differences in reduction in content versus function words, provides no specification of how content versus function word-hood was coded. Anttila et al. (2010), which studies the role of prosodic weight as measured by lexical stresses in the English dative alternation, also provides no list of words considered to be function words, even though function words crucially do not contribute lexical stress in the calculation of their dependent variable. For Shih et al.'s (to appear) study of rhythm in the English genitive alternation (and subsequent work: e.g., Grafmiller and Shih 2011), function word stress was taken directly from the Carnegie Mellon University Pronouncing Dic-

[39] Although, note that *–illion* seems to be the only truly productive cardinal number morpheme.

[40] To the best of the author's knowledge, no such master list exists.

tionary (Weide 1993) default, for which no coding descriptions are given to explain inconsistencies. For example, the possessive pronoun *their* is given in CMU as stressed; *her* is listed as unstressed first, with a second entry for a stressed variant; and *his* is listed as stressed first, with a second entry for an unstressed variant.[41]

In Temperley (2009), which studied rhythmic stress regularity in English, a list of function words was actually provided in an appendix; however, this list was researcher-specified, with little or no correction for homonyms. Selkirk (1984:352–354) provides lists of monosyllabic and polysyllabic function words, but, like others, this list is also by no means comprehensive: for example, *there*, *where*, *how*, and *which* are amongst the items missing from Selkirk's list that are otherwise typically considered to be function words. Bell et al. (2009:97) provide one of the more comprehensive categorizations of function words for their study, based on part-of-speech coding; however, as with most of these categorizations, Bell et al. arrived at their division of lexical and grammatical words based on prior assumptions about the division and not on word-by-word or part-of-speech-by-part-of-speech evidence.

It is clear that what is needed is a principled, evidentiary, and consistent method for distinguishing lexical and functional classifications. If we want to be able to appeal to a function and content word distinction in studies and theories of rhythm, then, to the extent possible, such a principled classification of the lexicon should be based in part on first understanding how content and function words behave in terms of stress and reduction.[42] Some headway on this front has been made with evidence from intonational prominence, as discussed in §2 (e.g., Altenberg 1987; Hirschberg 1993; Nenkova et al. 2007; a.o.). The practical goal of the current study is to offer such a principled division of the lexicon grounded in systematic phonetic evidence of

[41] To round out the possessive pronouns, *your* is listed as stressed, *my* as stressed, and *our* as first a two-syllable stressed-unstressed word with a stressed, monosyllabic variant.

[42] Do note, however, that I am in no way here claiming that phonetic evidence is the only definition of function versus content-word-hood acceptable. Clearly there are also syntactic and semantic factors to take into consideration. What I am arguing for, however, is a principled way of arriving at the classification of words, based on quantifiable evidence rather than researcher-biased intuitions. Such evidence need not obligatorily be phonetic in nature.

stress and reduction. Furthermore, to encourage replicability of the “master list” of content and function words proposed herein, standard Penn Treebank part-of-speech tags are used as the basis of word categorization.

4 Data

The current study utilizes data from the Buckeye Variation in Conversation corpus (ViC) (Pitt et al. 2007). The ViC corpus is a phonetically-transcribed collection of conversational American English, consisting of conversations with forty native American English speakers. Speakers were all from the Columbus, Ohio area. Conversations were facilitated on everyday topics by interviewers at the Ohio State University between 1999 and 2000. Each conversation lasted between thirty and sixty minutes, yielding about 300,000 words.

The ViC corpus includes phonetically-transcribed pronunciations, segment and word durations, and part-of-speech tags based on the Automatic Penn Treebank part-of-speech tags (Santorini 1990; Marcus et al. 1993). A random subset of the part-of-speech tags for monosyllabic words in the ViC corpus (n = 10,000) were hand-checked by the author for accuracy, and, if necessary, corrected. The dataset was then annotated for stress, dictionary pronunciation, and syllabic information, which were taken from the American English Unisyn lexicon (Fitt 2001). Words not available in the Unisyn lexicon were manually coded. Corpus-internal word frequencies, previous and following conditional bigram probabilities, speech rate, and phrase boundaries were calculated via Python scripts; these are further described in §7.1.

For this study, all monosyllabic words in the ViC corpus were extracted (n = 215,401). To clean the data, interjections and disfluencies (e.g., *mm*, *um* , *yeah*, *like*, *oh*) were excluded via hand-coding and automatic search for the part-of-speech tag ‘UH.’ The full set of monosyllabic words was further trimmed by removing all data in which the duration of the word and the average rate of speech surrounding the target word were three standard deviations away from the log mean (following Kuperman & Bresnan 2012; a.o.). The final dataset includes 206,858 tokens of monosyllabic words.

5 Quantifying reduction: phonetic correlates of stress

There are three primary phonetic correlates of stress: duration, amplitude, and pitch (examples modified from Hayes 1995:6):

1. DURATION: Stressed syllables are longer than reduced and unstressed syllables.

e.g., [ˈpɹ̩ː.mɪt] versus [pɹ̩.ˈmɪːt]

2. AMPLITUDE: Stressed syllables are louder than reduced and unstressed syllables.

e.g., ˈ*PERmit* versus *per*ˈ*MIT*

3. PITCH: Stressed syllables tend to carry pitch peaks whereas reduced and unstressed syllables do not.

e.g., *pérmit* versus *permít*
H H

It is acknowledged, however, that there is no simple direct phonetic correlate of stress. The difficulty lies in the fact that stress is parasitic, sharing phonetic properties with other parts of phonology. For example, pitch is shared with intonational systems and, in tonal languages, with contrastive tone. Duration is used as a cue to indicate (contrastive) vowel length, which can be independent from the stress system altogether.

Other diagnostics for stress include listener judgments and various phonological alternations that rely on stressedness (Hayes 1995). Due to the parasitic phonetic nature of stress, the phonological literature has primarily focused on phonological alternations in order to observe the presence of stress. In English, for example, we can diagnose unstressed syllables via vowel quality. Schwa [ə] only occurs in unstressed syllables, and in reduction, when stress shifts away from a given syllable, a full vowel will reduce to a schwa. In (43), the shift of primary stress from the first syllable of *párent* to the second syllable in *paréntal*, as caused by the *–al* suffix, reduces the vowel [ɛ] to [ə] in the first syllable of the word.

(43)	[ˈpɛ.ɹn̩t]	→	[pə.ˈɹɛn.tl̩]
	párent	→	*paréntal*

As with schwa reduction, segmental contrasts in general tend to be neutralized in reduced and unstressed syllables.

As noted in §2, the previous literature has relied largely on listener judgments of prosody and pitch accent when diagnosing stress (e.g., Altenberg 1987; Hirschberg 1993; Nenkova et al. 2007.). For example, Nenkova et al. (2007) utilizes a prosodically-annotated version of the Switchboard in which prominence was tagged by a human labeler based on the presence (by ear and visually on the pitch track) of pitch accent (Ostendorf et al. 2001). This methodology raises two issues. First, relying on listener-coded prosodic tagging can lead to inconsistencies, as noted in Ostendorf et al. (2001:2), who carried out the Switchboard prosody tagging. Hand-labeling also reduces the reliability of and potential for cross-corpus comparisons if different corpora are annotated by different listeners. Second, there exists a conceptual gap. On one hand, the empirical corpus work on prosodic differences between function and content words has focused primarily on pitch accent: presence of a pitch accent is equated with the presence of prominence or stress, broadly construed. On the other hand, the theoretical literature on stress differences between function and content words focuses primarily on reduction: presence of reduction is an indicator of the lack of stress. Not only is pitch merely one of at least three phonetic correlates of stress, but also the reduction literature usually points to duration and segmental differences (e.g., schwa reduction)—*not* pitch—as diagnostics of whether reduction has occurred.

Some work has been carried out on reduction from the perspective of duration (e.g., Bell et al. 2003, 2009) and segmental alternations (e.g., Johnson 2004a). The current study follows these works by presenting a large-scale comparison of both segmental reduction and duration in function and content words across a corpus of spontaneous speech. Each measure of reduction and how each is operationalized is presented in the following sections: segmental reduction in §5.1, and duration in §5.2. Section 5.3 presents a brief look at how segmental reduction and duration pattern according to the standard binary content versus function word division.

5.1 Segmental reduction

One phonetic cue to stress reduction is deviation in segmental material. As discussed above, full vowels will reduce to schwa when occurring in stressless environments. Schwa reduction is an example of a segmental alternation in reduced contexts that is phonologized in English, but a number of non-phonologized segmental alternations can accompany reduction in connected speech as well. For example, Johnson (2004a; following Stampe 1973), points out that *divinity* [də.ˈvɪ.nə.ti] can occur in the reduced variant [də.ˈvĩ.ə̪.ti] or, even further reduced, as [də.ˈvĩɪ̪̃]. While these "massive reductions," as Johnson calls them, can occur to a certain extent even on stressed syllables, the assumption is that the presence of stress inhibits reduction whereas the absence of stress encourages it. Thus, unstressed syllables should be less faithful to full, lexical forms—that is, more prone to differences in segmental qualities and to segmental deletions—than their stressed counterparts.

To quantify segmental reductions, I follow Johnson (2004a; 2004b) in using a measure of segmental deviation, which I call *phone distance*. The idea behind phone distance and segmental deviation is to have a calculation that compares the distance between actual, pronounced phones and their lexically-listed target phones: more deviance between the pronounced and target phones can be taken as a proxy of more reduction.

The measure of phone distance used here is taken from Johnson (2004b) and is based on the perceptual similarity between phones. Perceptual similarity was calculated (by Johnson 2004b) via a confusion matrix built on disagreements between transcribers of the ViC corpus. As reported in Raymond (2003), multiple transcribers coded the same subsections of the ViC corpus. The confusion matrix and perceptual similarity represents the disagreements between transcribers: for each pair of phones *i* and *j*, how likely one transcriber chooses *i* when another transcriber chooses *j*. The calculation for similarity is defined in (44). Then, the perceptual distance is taken to be the negative natural log of similarity (45).

(44) Similarity

$$S_{ij} = \frac{P_{ij} + P_{ij}}{P_{ii} + P_{jj}},$$

where S_{ij} = similarity between phones *i* and *j*;
P_{ij} = probability of disagreement between transcribers;
and $P_{ii,\,jj}$ = probability of agreement between transcribers.

(45) $r_{ij} = -ln(S_{ij})$,
where r_{ij} = perceptual distance between phones *i* and *j*.

Using this measure r_{ij} of perceptual distance between two phones, the *phone distance* for a given word was calculated by summing the perceptual distance between every pair of phones in the pronounced word and the target, dictionary-listed, citation form of the word. The resulting sum was normalized onto a scale of 0 to 2. All phone distances used for the current study were taken directly from Johnson's data.[43] For more details on these calculations, see Johnson (2004b:3–4).

On the phone distance scale (46), **0** represents no deviation between the citation form and the pronounced form. A measure of **2** represents the maximum deviation between citation and pronounced forms.

[43] Acknowledgements to Keith Johnson for sharing his phone distance values.

(46) *Phone distance measure*:

0 ⟶ 2

no reduction ⟶ more reduction

Example tokens from the ViC corpus of the pronunciation of *things* and their phone distance values are given in (47). A period [.] in the transcription represents a segment in the citation form that has no counterpart in the pronounced form.

(47)

Word	Citation form	Pronounced form	Phone distance
things	[θɪŋz]	[θɪŋz]	0
		[θɪŋs]	0.123
		[θɛŋz]	0.146
		[θɪ̃.z]	0.44
		[θeɪŋs]	0.484
		[sɪŋ.]	0.563
		[sɪ̃.ʒ]	0.911
		[h.ŋs]	1.05

The overall distribution of phone distance measures in the dataset of monosyllabic ViC words is shown in Figure 12. A large portion of the data (44.02%, n = 91,067) exhibits perfect correspondence between pronounced and citation forms (i.e., phone distance = 0), perhaps due to a pressure to pronounce words faithfully. Very few tokens incur a phone distance value of 2 (n = 11). All words in this latter set, except one (*know*), are tokens of the function words *a* and *I*. Seventy-five percent of the data falls below a phone distance value of 0.468.

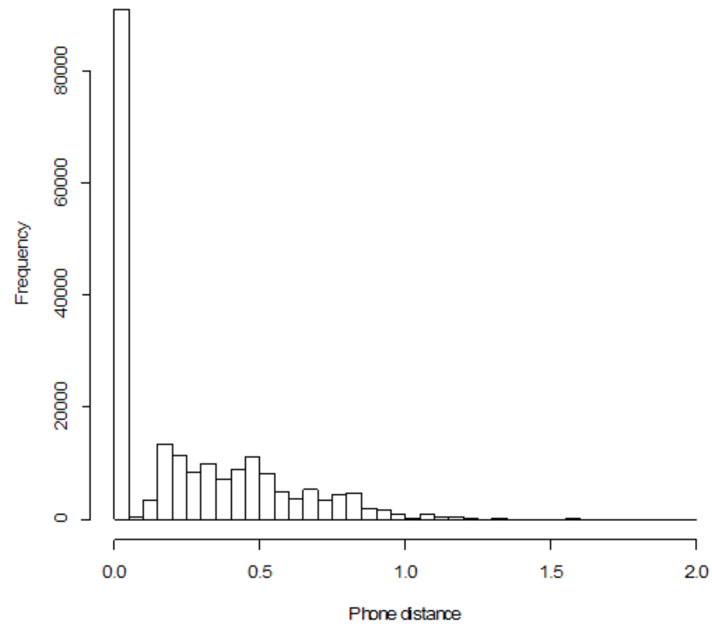

Figure 12. Frequencies of phone distance values for monosyllabic words in the ViC corpus

One caveat to note is that the phone distance measure used here only takes into account perceptual similarity and has nothing to say about similarities in production. Furthermore, the literature on confusability and similarity matrices is vast (e.g., Mielke 2012); however, I will assume Johnson's distance measures here and leave the question of alternative measures to future work.

5.2 Duration

Another phonetic, quantifiable cue to stress is duration. In perception, an increase in syllable duration increases the likelihood the syllable will be perceived as stressed, more so than any increase in intensity (Fry 1955). In production, durational differences have been shown to be more closely linked to the difference between stressed and unstressed syllables, whereas differences in pitch denote accentual differences (Beckman & Edwards 1994). The expectation, then, is that, with all else being equal, stressed syllables will be longer than unstressed ones. This pattern should correlate with content words, which are lexically stressed, being longer than function words. For example, *can* as a noun should be longer than *can* as a modal auxiliary. In the ViC corpus, *can* in the phrase *a can of beans* has a duration of 0.295 seconds. In contrast, *can* in the phrase *I can afford to hang on* exhibits a shorter duration of 0.141 seconds. Both are pronounced [kɪ̃r̃].

As with segmental reduction (§5.1), I use here a measure of duration, which I call *duration distance*, that incorporates a pronounced word's deviation from a target, full form, operating under the assumption that reduction from a citation form will cause shorter durations. Estimated citation durations of each word were taken by summing, for each phone in the dictionary-listed pronunciation of the word, the average duration of that phone across its every appearance in the ViC corpus (48).

(48) Estimated citation duration for a given word *w*

Let $C(w) = (i_1, i_2, i_3, \ldots, i_n)$,

$$T_e = \sum_{k=1}^{n} \bar{T}(i_k),$$

where $C(w)$ = citation, dictionary-listed form of *w*;
$i_k = k^{th}$ phone in $C(w)$;
n = number of individual phones i in $C(w)$;
$\bar{T}(i)$ = average duration of phone i in the ViC corpus;
and T_e = estimated duration of $C(w)$.

Duration distance is calculated as a ratio of the duration of the pronounced instance and an estimated citation duration of the word. The ratio is then normalized by the number of phones in both the dictionary and transcribed pronunciations (49):

(49) Duration distance

$$D_r = \frac{T_e - T(w_r)}{n + n_r},$$

where $w_r = r^{th}$ pronounced instance of w;
D_r = duration distance of w_r;
$T(w_r)$ = duration of w_r;
and n_r = number of individual phones in w_r.

In their study of reduction and phonological neighborhood density in the ViC corpus, Gahl et al. (2012) use a similar methodology as the one used in the current study; however, they hold the duration of the target pronunciation as a control variable whereas it is incorporated directly into the duration measure here.

A resulting scale of duration distance is given in (50). As with the phone distance measure, a more positive number for duration distance indicates more reduction, and a more negative number indicates less reduction.

(50) *Duration distance measure:*

$$\underset{\text{less reduction}}{-} \longleftrightarrow \underset{\text{more reduction}}{+}$$

To return to the brief example of *can* discussed above, the estimated citation duration for the word *can* is 0.284 seconds. Given the estimated citation duration, the noun instance of *can* (i.e., *a can of beans*) has a duration distance value of -0.00165. In contrast, the modal auxiliary *can* in *I can afford to hang on* incurs a duration distance value of 0.0235. This difference between the two instances of *can* suggests that there is less reduction for the *can* that is a noun and lexically stressed as compared to the function word *can.*

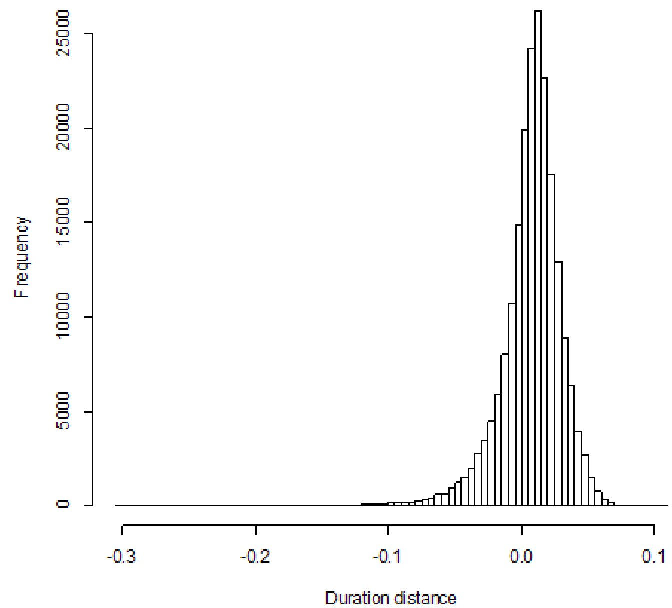

Figure 13. Frequencies of duration distance values for monosyllabic words in the ViC corpus

The overall distribution of duration distance for monosyllabic words in the ViC corpus is shown in the histogram in Figure 13. A majority of the data exhibits shorter pronunciation durations from their corresponding estimated citation durations (71.64%, n = 148,198). Because the citation durations used are only estimates, duration distance values smaller than 0 indicate longer durations than expected. However, these longer-than-expected durations may be due to underestimation in the pronunciation of citation forms when average phone durations were taken from the overall corpus.

5.3 A brief look at reduction in content versus function word classes

This section explores the patterns of segmental reduction and duration in content and function word classes. Content and function word classes were coded based largely on the Penn Treebank part-of-speech tags available in the ViC corpus, along with some lexical fine-tuning for verb forms, as shown in Table 6.

Table 6. Binary division: Content versus function words

Content words		Function words	
Part-of-speech	Treebank tag	Part-of-speech	Treebank tag
Nouns	NN, NNS, NNP, NNPS	Pronouns (including demonstrative, quantifying, personal, relative, possessive, and <u>*wh*</u>-pronouns)	PRP, PRP$, WP, WP$
Lexical verbs[44]	VB, VBP, VBZ, VBD, VBG, VBN (except non-lexical verbs)	Non-lexical verbs (all forms of *be, do, have*)	VB, VBP, VBZ, VBD, VBG, VBN
Adjectives	JJ, JJR, JJS[45]	Modal verbs	MD
Adverbs	RB, RBR, RBS	*Wh*-adverbs	WRB
		Articles	DT
		Conjunctions	CC, IN
		Prepositions	IN, TO
		Participles	RP
		Wh-determiners	WDT
		Predeterminers	PDT
		Existential *there*	EX
		Cardinal numbers	CD
		Contractions	DT_VBZ, EX_VBZ, MD_RB, NNS_VBZ, PRP_MD, PRP_VBP, PRP_VBZ, VBP_PRP, VBP_RB, WP_VBZ

[44] There are actually no gerund, present participle lexical or non-lexical verbs (e.g., *being*, *doing*; tagged 'VBG') in the dataset because these verbs are never monosyllabic.

[45] The comparative and superlative adjectives and adverbs (JJR, JJS, RBR, RBS) are limited to the following set: *more*, *most*, *worse*, *worst*, *best*, *less*, *least*. Regular or derived comparatives and superlatives are not monosyllabic (e.g., *brighter*, *brightest*), and therefore were not included in this dataset.

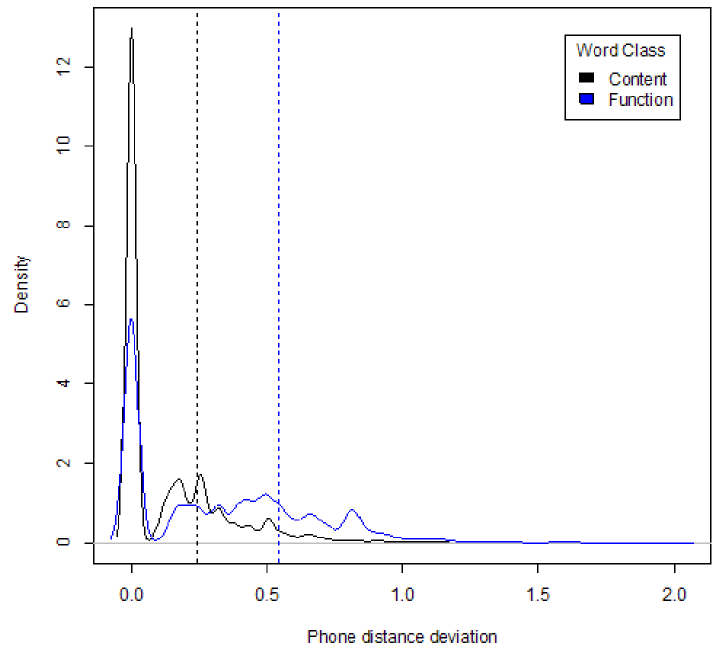

Figure 14. Probability densities for Phone Distance in Content versus Function words. (Dotted vertical lines denote third quartile of each probability distribution.)

Under the assumptions that content and function words feature differences in stress properties, we expect that function words should exhibit more segmental reduction and shorter durations than content words. The data also demonstrates that finer-grained differences, as described by Altenberg (1987) and others (e.g., Hirschberg 1993) hold: just as the data shows differences in reduction between content and function words, there are similar differences in reduction within the large family of function words.

Comparing content and function words in the coarse, binary division given in Table 6, there is a significant difference between the phone distance value distributions for the two classes of words (Kolmogorov-Smirnov: $D = 0.326$, $p < 0.0001$). As shown in the density plot of phone distances by content and function words in Figure 14, monosyllabic content words tend to have a lower phone distance value than monosyllabic function words—that is, content words are more likely to be pronounced closer to citation forms than monosyllabic function words. This can also be observed by examining the quartiles of the probability densities. The dotted vertical lines in the

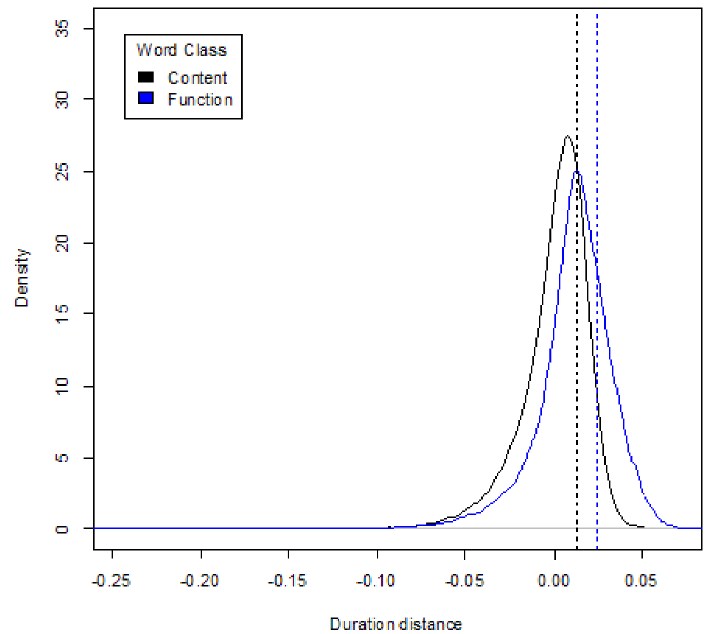

Figure 15. Probability densities for Duration Distance in Content versus Function words. (Dotted vertical lines denote third quartile of each probability distribution.)

plot denote the third quartile of each probability density function. As expected, the third quartile of function words is greater than that of content words—that is, seventy-five percent of content words fall below a lower phone distance value (*phone distance* = 0.243) than seventy-five percent of function words (*phone distance* = 0.54). This indicates that more function words exhibit more segmental reduction than content words.

Similar results obtain for duration distance, as shown in Figure 15, with significantly different distributions for function and content words (D = 0.2463, p < 0.0001). The density plots for duration distance by content and function words peaks further to the left for content words than for function words, indicating that function words are generally shorter in duration than content words. The third quartiles also demonstrate the same pattern: seventy-five percent of the content words have a duration distance value less than 0.01342 while the upward bound of duration distance values for seventy-five percent of function words is 0.0246.

The boxplots in Figure 16 also make evident the significant differences in reduction between content and function words: content words generally have less seg-

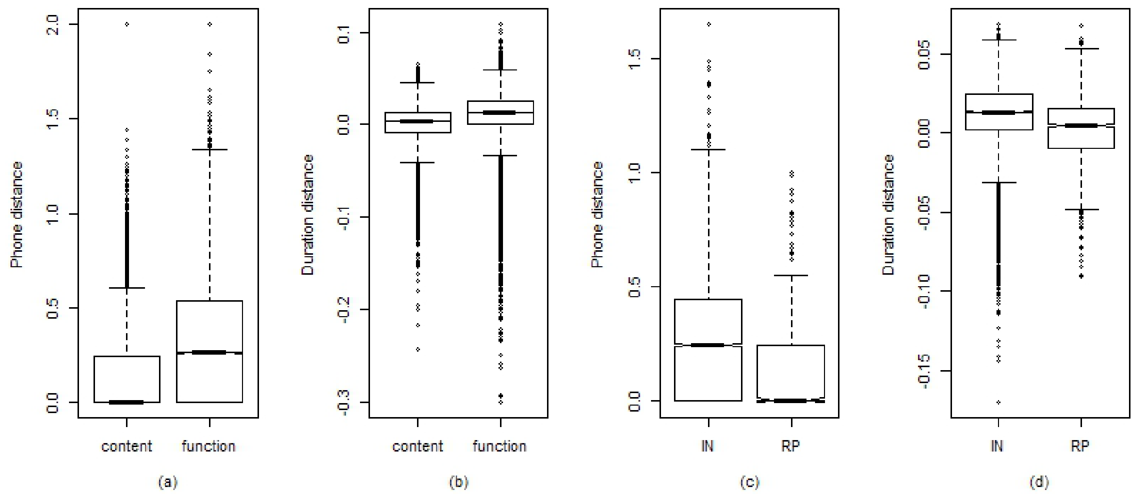

Figure 16. (Left to right) (a) Boxplot of phone distance values by content and function words; (b) Boxplot of duration distance values by content and function words; (c) Boxplot of phone distance values by prepositions (IN) and particles (RP); (d) Boxplot of duration distance values by prepositions and particles.

mental reduction (i.e., smaller phone distance values; Figure 16a) and longer durations (i.e., smaller duration distance values; Figure 16b) than function words.

Looking more closely at the patterns of reduction, we see that within function words themselves, there are significant differences between how often reduction occurs. For example, prepositions are much more likely to exhibit segmental reduction and shorter durations from their citations forms than particles are, despite the fact that many prepositions and particles are homophonous:

(51) a. Preposition (IN): *there's really no aerospace things* ***in*** *Columbus*
phone distance in ViC = 0.244

b. Particle (RP): *sometimes we have to bring* ***in*** *a couple of instrumentalist*
phone distance in ViC = 0

The boxplots in Figure 16c–d shows significantly different means between the distributions of prepositions (IN) and particles (RP) for both phone distance and duration distance measures. Note that prepositions have, on average, more reduction than particles: their difference is similar to the difference between function and content words in Figure 16a–b.

Figures 17 and 18 provide comparisons of phone distance and duration distance by part-of-speech. As expected, words traditionally classed as “content words”—that is, nouns and verbs—tend to have smaller phone distance and duration distance values. Forms featuring contractions, as marked by compounding of part-of-speech tags (e.g., DT_VBZ; see Table 119 for a full list), tend to exhibit greater phone distance and duration distance values than content words: this behavior is as expected given the nature of contractions, and for the purposes of the rest of the current study, they will be set aside from monomorphemic function words.

Within function words, there are differences as well. Determiners (DT) have a significantly greater mean phone distance than other function words, including prepositions, modal auxiliaries, and pronouns. Prepositions (IN) have a greater mean phone distance than modals, pronouns, and particles, as previously demonstrated. While the patterns of segmental reduction and duration are generally quite similar, there are some divergencies: for example, pronouns (PRP) and determiners have much greater means for phone distance as compared to the other function word parts-of-speech, but they exhibit less duration distance. Content words themselves are not an entirely homogeneous set, either, but their variance in means is not as large as that of the set of words standardly considered to be in the functional category. These differences within the standard content and function word categories suggest that stress and prosodification may rely on more than a simple binary split in the lexicon. The following section (§6) builds on this observation with a systematic approach to natural groupings of the data based on differences in segmental reduction and duration.

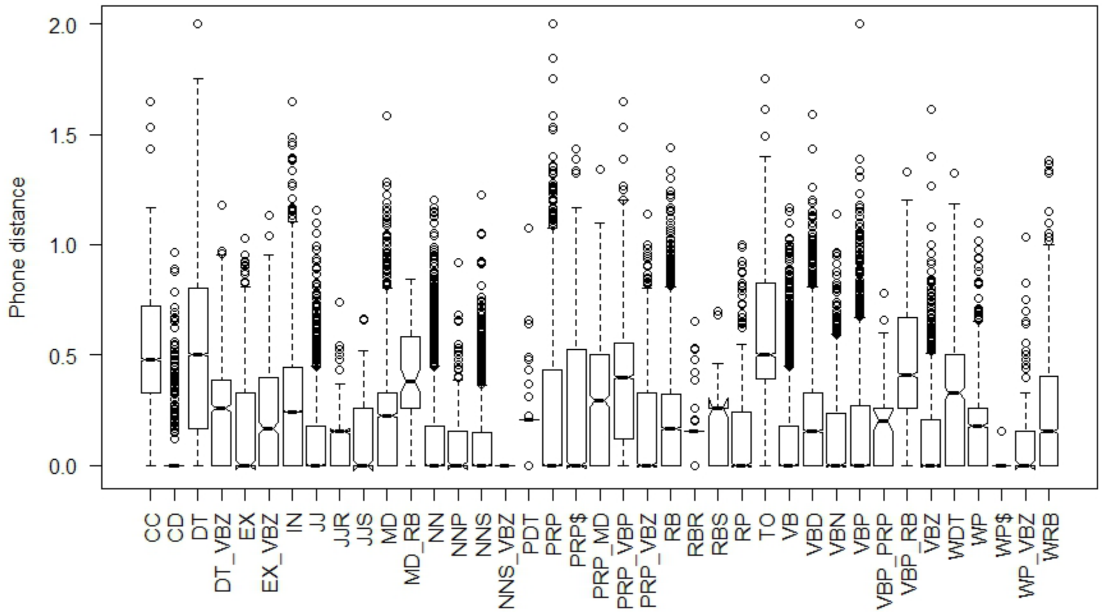

Figure 17. Box plot of phone distance values by part-of-speech.

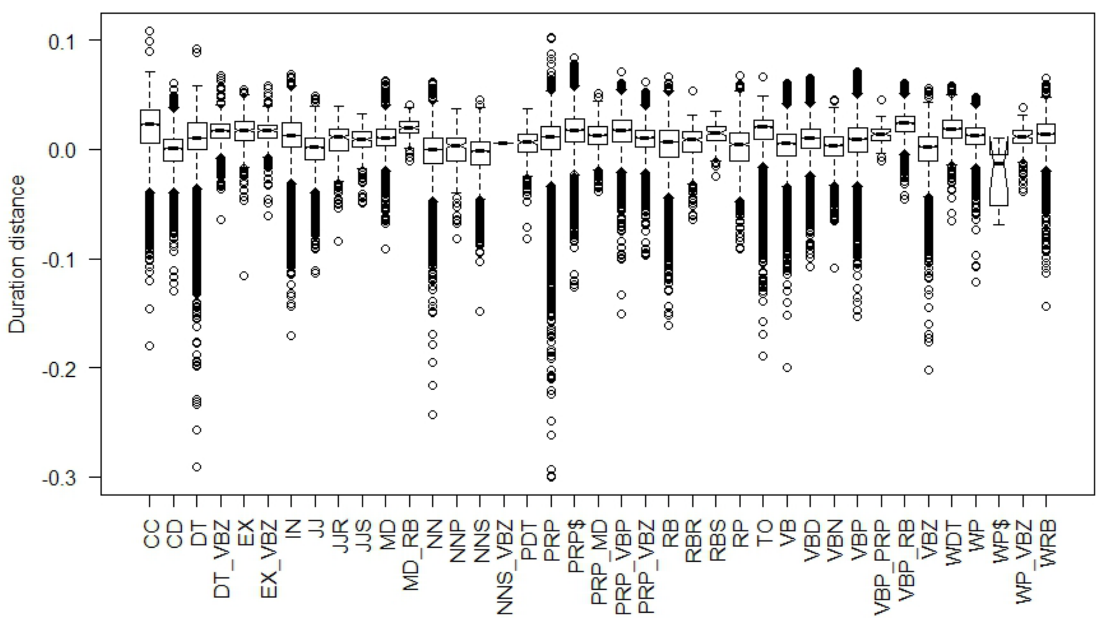

Figure 18. Box plot of duration distance values by part-of-speech

6 Towards a bottom-up categorization of words

One main question that follows from observing the significant differences in reduction between certain parts-of-speech is to what extent words form larger groups that pattern together in reduction and stress. Previous work has varied in their approaches to this question, either—most commonly—drawing large divisions across words and parts of speech or making no assertion of grouping and relying solely on narrow part-of-speech or lexical distinctions (e.g., Nenkova et al. 2007). Here, I take a bottom-up methodology to investigating this question using hierarchical cluster analysis to find groupings within the data, given the measurements of segmental reduction and duration. Categories that are uncovered from the data are taken to be meaningful if they [1] improve the power of the analysis of reduction patterns without overfitting (i.e., being too lexically-specific and thus un-generalizable to new data); [2] demonstrate independent differences as groups with respect to factors (e.g., frequency, rhythm, structural position) that should distinguish between content and function word classes; and [3] concord with theoretical expectations and/or previous findings. The current section introduces the hierarchical cluster analysis and results, and the diagnostics of these results are taken up in §§8–9.

6.1 Examining segmental and duration reduction: Third quartiles

Words—or classes of words—that exhibit more reduction should not only have greater distributional ranges in the phone and duration distance measures but also have more data mass located at higher values of these measures. These characteristics indicate that these words exhibit higher levels of reduction and have greater rates of doing so, respectively. To examine the patterns of segmental reduction and duration together, it is necessary to have a measure that captures both of these pieces of information. The standard methods of comparing means and medians does not provide adequate information about ranges *and* data mass at higher ranges. Thus, here, third quartile values of phone distance and duration distance are used instead. The third quartile (Q_3) of a data distribution splits the highest twenty-five percent of the data from the lower seventy-fifth percentile, meaning that seventy-five percent of the data occurs lower than the Q_3 cut-off value. In Figures 17 and 18 above, we saw that the third quartiles of the density estimations reflected expected differences in the distribution of the data: a higher Q_3 indicates more reduction amongst that given set of data (e.g., function words) as compared to another with a lower Q_3 value (e.g., content words). As opposed to comparing means, comparing third quartiles allows us to capture some information about the upper range of the distribution of reduction measurements, beyond what other measurements like means or medians can tell us—and particularly when dealing with non-normal or unknown distributions, as we are here.

Quartiles for phone distance and duration distance were calculated for each Penn Treebank part-of-speech tag in the data, using median-unbiased Type 8 quantile estimation in R (Hyndman & Fan 1996). Resulting Q_3 cut-offs are given in Table 7.

Table 7. Third quartile (Q_3) phone distance and duration distance values by part-of-speech (using unique Penn Treebank part-of-speech tags)

Part-of-speech	**3rd Quartile Phone Distance**	**3rd Quartile Duration Distance**
Function words		
Coordinating conjunction	0.721	0.0358771
Cardinal number	0	0.009173542
Determiner	0.8	0.0241761
Existential there	0.325	0.0253255
Preposition	0.441	0.0245999
Preposition *to*	0.824	0.0270903
Modal auxiliary	0.325	0.01825989
Predeterminer	0.208	0.01405497
Pronoun	0.433	0.0214673
Possessive pronoun	0.5296667	0.0280807
Particle	0.24	0.01539
Wh determiner	0.502	0.0264461
Wh pronoun	0.26	0.01963513
Possessive *Wh* pronoun	0	-0.003918867

Wh adverb	0.403	0.0230041
Content words		
Adjective	0.18	0.01108987
Comparative adjective	0.153	0.0185058
Superlative adjective	0.26	0.01611823
Noun, singular	0.18	0.01019105
Proper noun, singular	0.156	0.01106357
Noun, plural	0.146	0.007351467
Adverb	0.324	0.01714268
Comparative adverb	0.153	0.01662243
Superlative adverb	0.26	0.02082704
Verb	0.18	0.0135354
Past tense verb	0.325	0.01822855
Past participle	0.237	0.01174929
Present verb (not 3rd person)	0.268	0.01942262
Present verb (3rd ps.)	0.205	0.0118527
Contractions		
Determiner_Verb (pres.)	0.383	0.02293542
*there*_Verb (pres.)	0.4	0.02260518
Modal_Adverb	0.584	0.0252345
Noun_Verb (pres.)	0	0.0063035
Possessive pronoun_Modal	0.5	0.020942
Possessive pronoun_Verb (pres., not 3rd ps.)	0.554	0.02653597
Possessive pronoun_Verb (pres., 3rd ps.)	0.325	0.01758138
Verb_Pronoun	0.26	0.01864025
Verb_Adverb	0.6683333	0.0305571
Wh pronoun_Verb (present tense)	0.154	0.01766236

Eyeballing the Q_3 cut-offs in Table 7, the difference between content and function words, as was graphically noted in Figures 17 and 18 above, is evident. Figure 19 plots the standardized Q_3 values of phone distance against the standardized Q_3 values of duration distance, with the standard content and function divide represented in blue and red, respectively.[46] With the exception of a few parts-of-speech, function words generally exhibit a higher range of phone and duration distance Q_3 values than content words, suggesting that more reduction is present in function words than in content words. For phone distance, most Q_3 cut-offs for function word parts-of-speech fall roughly within a 0.2 to 0.8 range while for most content words, phone distance Q_3 cut-offs range between 0.14 and 0.32. Duration distance third quartiles are similar: the ma-

[46] Contractions are temporarily excluded from Figure 19 for exposition and clarity purposes.

jority of function word Q_3 values range between 0.014 to 0.036, as compared to a majority range of 0.007 to 0.02 for content words.

Notable exceptions include possessive *wh* pronouns and cardinal numbers, which are outliers with low Q_3 values, exhibiting very little reduction from their citation forms (see bottom left corner of Figure). The lower-than-expected values for possessive *wh* pronouns could stem from data sparseness, with only nine instances of *whose* in the dataset. Cardinal numbers, however, do not suffer from data sparseness (n = 2548); thus, their low phone distance and duration distance values could reflect an actual lack of reduction for these words. This pattern in cardinal numbers suggests that, despite being a somewhat more closed class lexically when compared to other nouns (though cf. recently-coined numerals such as *gazillion*, *bajillion*), numbers legitimately belong under the umbrella of content words, according to the stress data. Contractions tend to fall in a midrange of the phone distance and duration distance measures between content words and the function word part-of-speech classes that have the highest Q_3 values. The single exception is the noun+verb contraction, which only has one token in the dataset (*wife's*), and thus cannot be meaningfully assessed.

From Table 7 and Figure 19, we can also observe that the function word class has a much greater range of Q_3 values than the content word class. For phone distance, the range is approximately 0.54 for function words, excluding cardinal numbers and possessive *wh* pronouns. In comparison, the phone distance Q_3 range for content words is much more narrow: 0.179. Likewise, for duration distance, the range for function words is 0.0218, as compared to 0.0135 for content words. The differences show that, though there is also some variation for reduction within content words, the range of variation for reduction in function words is greater. Some function word part-of-speech categories share similar Q_3 cut-offs as content words—for instance, predeterminers and particles—while others, like determiners, reveal larger differences. This pattern follows Altenberg's (1987) observations of function and content word differences: by part-of-speech type, content words vary less in their prosodic potentials than function words. Altenberg took this observation as motivation for his scalar approach to word categorization over a standard binary division, which failed to capture the

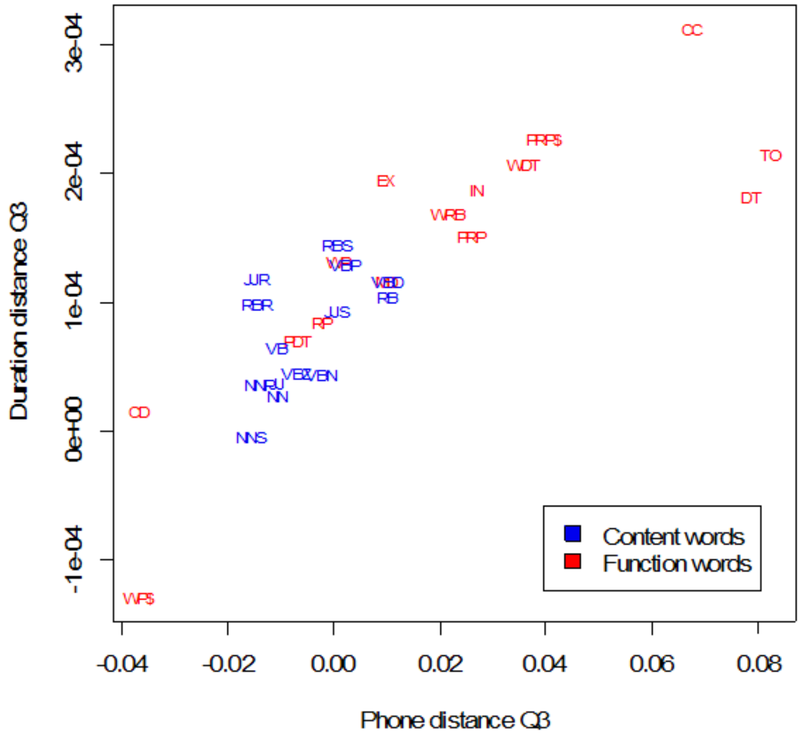

Figure 19. Third quantile (Q3) values of phone distance and duration distance by standard content versus function word divisions.

wide spread of function word prosodification—as opposed to a narrower spread for content words—in his study.

From the raw numbers, it is easy to see the larger differences between content and function words. But, the main questions at hand are whether subgroups of words pattern together in terms of their potentials for stress and reduction and, if so, what these subgroups are. To address these issues, I now turn to a hierarchical cluster analysis of the reduction patterns.

6.2 Hierarchical cluster analysis

Hierarchical cluster analysis is an unsupervised, agglomerative, bottom-up method of discovering groupings in the data based on similarities and distances between each observation and every other observation—in this case, between each part-of-speech and every other part-of-speech—on the basis of the independent variables—segmental reduction and duration. Hierarchical cluster analysis is advantageous for use in the current problem because there is no need to pre-specify the number of groups that the model should find, as is necessary with other cluster methods. Because we do not need to make any *a priori* assumptions about the final number of word categories and divisions, this methodology ameliorates preconceived research subjectivity and biases and allows us to more objectively examine what the patterns in the data are showing.

Here, the analysis is based on the phone distance and duration distance Q_3 cut-off values by parts-of-speech in dataset. It should be noted that, in theory, hierarchical cluster analysis could be carried out on pairwise distance comparisons between tokens; however, such an approach is unfeasible given severe technological limitations (i.e., the memory needed to store the distance matrix of every token compared with every other token for the ViC dataset). By-word comparisons are also possible, using the Q_3 cut-off value of each word's probability density, but the number of tokens varies widely by word, with many amounting to one token per type. Aggregating words by slightly larger categories such as parts-of-speech provides results that not only avoid potential data sparseness problems but also are graphically interpretable (e.g., see Figure 19). Moreover, parts-of-speech make for a sensible jumping-off point and point of comparison for the current analysis, given that they have served as the basis of many previous content-function word studies.

For the analysis, phone distance and duration distance measures were centered and standardized by subtracting their means and dividing by two standard deviations (Gelman 2008; a.o.). Q_3 cut-off values for the standardized phone distance and duration distance measures for each part-of-speech tag were taken.[47] Based on these values, a Euclidean distance matrix was calculated using the `dist` function. The resulting

[47] See also the unstandardized values in Table 7. Standardization does not affect the discussion or interpretation given above; its primary purpose is to ensure that phone distance and duration distance are given equal weight in the cluster analysis.

40×40 matrix contains the distance between each part-of-speech tag and every other part-of-speech tag in the dataset.

The hierarchical cluster analysis used herein (`hclust, method='Ward'`) works in the following way. Beginning bottom-up, each part-of-speech is a single "cluster." Using Ward's method (Ward 1963), pairs of clusters are formed by minimizing the increase of total within-cluster variance at each step: that is, amongst the possible pairs, the pair that produces the minimal amount of increase in the variance within the cluster, as calculated by sum-of-squares, is chosen. In this way, hierarchical cluster analyses are the bottom-up version of top-down partitioning models such as classification and regression trees (see Chapter 4). The algorithm continues until a complete cluster tree is produced and no more pairs are available for grouping. Ward's method was used because it tends to produce equally-sized, tightly-clustered spherical groups, due to its reliance on the sum-of-squares statistic. This means that the groups found contain members that are all maximally similar in both phone distance and duration distance. A complete-linkage model, which is instead based on comparisons of maximum distances between single members of clusters, was also generated with minimal differences in results and is included in Appendix B.

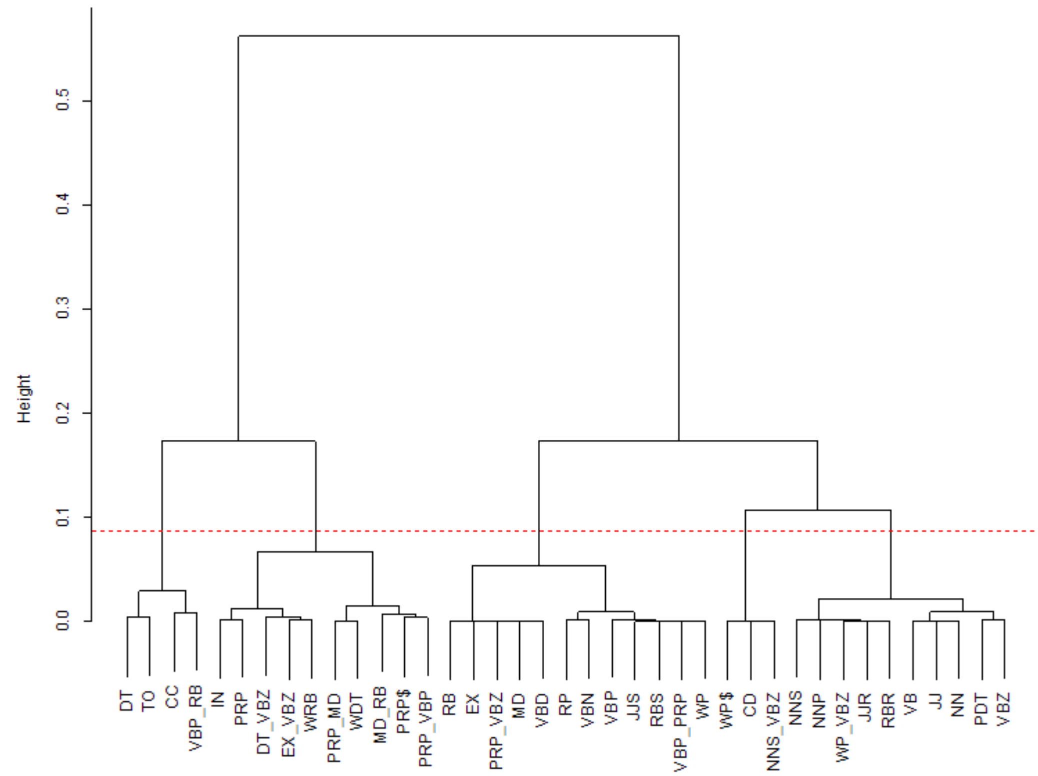

Figure 20. Hierarchical cluster dendrogram: Part-of-speech clusters by phone distance and duration distance. Dashed horizontal line represents a five-cluster solution.

The results of the hierarchical cluster analysis are given in Figure 20. The height of the lines in the dendrogram (*y*-axis) denotes how dissimilar the two component clusters are when they were merged: taller lines indicate greater distances between the clusters. The content versus function word division is clear from the hierarchical model. The left branch of the dendrogram is made up entirely of parts-of-speech standardly regarded as function words: determiners, conjunctions, prepositions, *wh* adverbs, and *wh* determiners. The right branch includes lexical nouns, verbs, adjectives and adverbs. In the right branch, however, we also find grammatical parts-of-speech: modals, particles, predeterminers, existential *there*, cardinal numbers—if we consider these to be function words—, and the comparative and superlative adverbs and adjectives (see fn. 45).

The cluster model also uncovers smaller groupings under the largest split based on the phonetic measures of reduction. Given the height of the lines, there are four or five major clusters of words: these groups can be seen in Figure 20 by tracing a horizontal line at about 0.1 on the *y*-axis and examining the nodes belonging to each of the vertical lines that the horizontal trace crosses. Membership of these five clusters are listed in Table 8 below. A four cluster grouping results in the lumping together of Clusters 1 and 2. While there are even smaller clusters underneath the five major subgroups listed in Table 8 (e.g., a cluster that includes pronouns versus a separate cluster that includes possessive pronouns), the comparative shortness of their lines before merging suggest that these smaller clusters may not be as significantly distinct as the larger clusters; thus, I will start here by examining a four/five-cluster solution.

Table 8. Five-cluster hierarchical cluster result groups by part-of-speech

Cluster	Parts-of-speech
Cluster 1	cardinal number, possessive *wh* pronoun, noun_verb
Cluster 2	adjective, comparative adjective, noun (singular and plural), proper noun, predeterminer, comparative adverb, verb, present verb (3rd ps.), *wh* pronoun_verb (pres., 3rd ps.)
Cluster 3	existential *there*, superlative adjective, modal auxiliary, adverb, superlative adverb, particle, past participle verb, past tense verb, present verb (not 3rd ps.), *wh* pronoun, pronoun_verb (pres., 3rd ps.), verb_pronoun
Cluster 4	preposition, pronoun, pronoun (possessive), *wh* determiner, *wh* adverb, determiner_verb (pres.), *there*_verb (pres.), modal_adverb, pronoun_modal, pronoun_verb (pres., not 3rd ps.)
Cluster 5	coordinating conjunction, determiner, preposition *to*, verb_adverb

The clusters are also plotted by color in Figure 21. As can be seen graphically, the hierarchical clustering model has selected groups that are tightly and centrally clustered according to how much segmental and duration reduction they exhibit. On one end of the scale of reduction, we see words that are standardly in a binary division classified as content words in Clusters 1 and 2, including nouns, adjectives, and verbs. Amongst all of the parts-of-speech examined, members of these two clusters tend to reduce the least in pronunciation from their citation forms. There are also a very limited number of parts-of-speech that are standardly not considered part of the content word class that

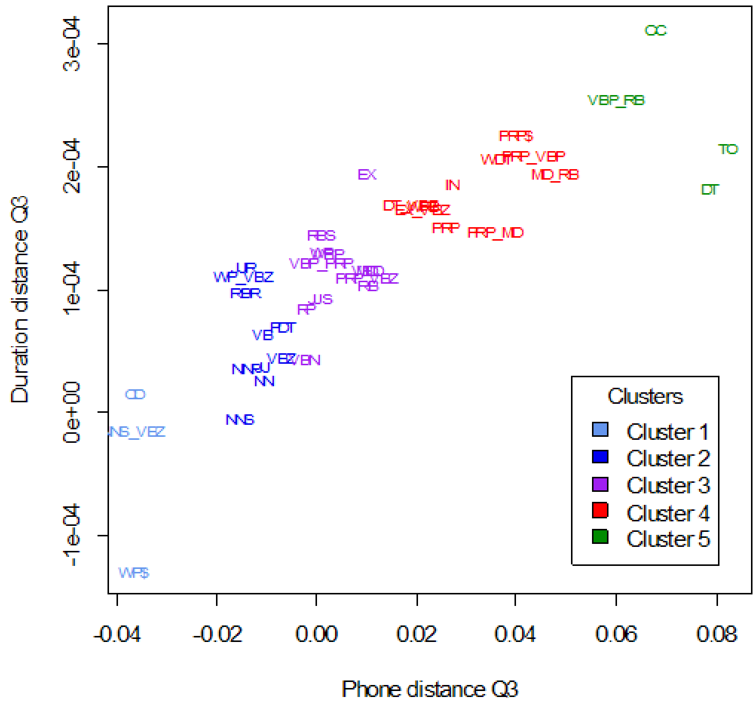

Figure 21. Third quantile (Q_3) values of phone distance and duration distance by hierarchical model part-of-speech clusters.

make it into Clusters 1 and 2: cardinal numbers, the possessive *wh* pronoun, and pre-determiners. Predeterminers (n = 308) include the items *all*, *such*, and *half*.

More towards the center of the scale of reduction but still located within the right "content word" branch of the hierarchical model is Cluster 3. This cluster consists of typically-considered content words such as inflected verb forms (e.g., past tense, past participle) and adverbs. The result that adverbs fall in Cluster 3 instead of in Clusters 1 and 2 , which contain most nouns, verbs, and adjectives, is interesting to note given that the literature is often inconsistent in its inclusion of adverbs in the lexical class. Also in Cluster 3, we find more parts-of-speech that are typically considered function words than we do in Clusters 1 and 2: existential *there*, modal auxiliaries, particles, *wh* pronouns, and superlatives.[48] Given the similarities of grammatical and

[48] Given the limited nature of the superlative adjectives and adverbs in the dataset (see fn. 45), it is reasonable to include them in this set of function word-like items.

semantic function and phrasal distribution between modal auxiliaries and verbs, it is unsurprising to see that they also behave similarly in terms of reduction patterns.

The fourth and fifth clusters, located in the large left branch of the hierarchical model in Figure 20, contain only parts-of-speech that are standardly considered to be function words. That is, unlike the appearance of function words in the other three clusters, no nouns, adjectives, verbs, or adverbs of any form occur in these two clusters (ignoring, for the time being, contractions). Cluster 5, which includes determiners, conjunctions, and the preposition *to*, represents parts-of-speech that exhibit the most reduction from their expected citation forms, both in terms of how much they reduce and, proportionally, how often reduction occurs. Cluster 4, which contains other parts-of-speech such as prepositions, pronouns, and most *wh* words, demonstrate smaller phone and duration distance Q_3 values—that is, generally less reduction—than their Cluster 5 counterparts.

The hierarchical cluster analysis reveals distinct word categories that do not entirely align with the standard binary classification of content and function words. On one hand, we see that the cluster analysis is able to—with little to no supervision—uncover distinct content and function word categories that for the most part concur with previous assumptions about this division. On the other hand, there are clear subgroups that are also uncovered. Some of these subgroups contain a mix of words that would otherwise be classed separately in a traditional division of the lexicon but here behave similarly in terms of segmental and durational reduction. If it is the case that we want to base assumptions about lexical stress and the structure of lexical categories on patterns of reduction in speech, then the results of the hierarchical cluster analysis suggest that the standard, strict binary division between content and function words must be revisited, with subtler groupings and distinctions in mind.

6.3 Comparisons with previous, non-binary word categorizations

As discussed above, the results of the hierarchical cluster analysis in §6.2 show not only a large categorical split between content and function words but also refined subgroups that include a combination of both lexical and grammatical words. Such results align with conclusions from previous work, as summarized in §2. For example, Hirschberg (1993:322) notes that while a simple binary content versus function word division is able to predict the placement of pitch accent with about 68% accuracy, a four-way category split (open, closed_accented, closed_deaccented, closed_cliticized) improved the accuracy of pitch accent prediction by nine percent. In this current subsection, the results of the hierarchical cluster analysis above are compared with previously proposed non-binary word categorizations that take into account parts-of-speech and/or lexical specifications. In particular, I focus here on comparing the clusters with Altenberg's (1987) word classes, which is based on prosodic potential measurements most closely aligned with the measurements of reducibility used here. Other non-binary word categorizations (e.g., Hirschberg 1993; a.o.) were based primarily on pitch accent evidence, which does not have a direct, one-to-one correlation with stress (e.g., see discussion in Hirschberg 1993: 307), and thus may give rise to significantly different word categorization results. Possible refinements to the cluster analysis based on lexical specifications in comparison with Altenberg (1987) are proposed.

The results of the cluster analysis on reduction in the ViC corpus (e.g., Figure 20) are quite similar to Altenberg's findings (see Table 4), though notable differences also abound. The immediately evident similarity is the larger spread of "function" word behavior: that is, some parts-of-speech standardly classified as function words exhibit similar levels of reducibility as content words do while others are clearly quite different, which higher levels of reducibility. Looking at specific similarities, the cluster analysis and Altenberg's classifications both find that determiners are the most reducible words, more so than many of the other function words. Prepositions in both analyses sit at the middle of the range of reducibility. Both analyses also find that particles, predeterminers, and cardinals demonstrate low levels of reduction.

The majority of the differences between the cluster analysis results and Altenberg's classifications are summarized as follows. In Altenberg's results, existential *there* demonstrates high levels of reducibility similar to determiners and possessive pronouns, but the same item in the ViC corpus patterns with more mid-range categories including adverbs and modal auxiliaries. Altenberg places modal auxiliaries in a class with grammatical words where the cluster analysis finds that they pattern with inflected forms of verbs in the content word category. Altenberg also finds that conjunctions exhibit less reduction relative to pronouns and determiners; but here we see that coordinating conjunctions pattern with determiners by demonstrating more reduction than the other part-of-speech categories. The categorization of *wh* adverbs is also quite different: for Altenberg, *wh* adverbs exhibit very little reduction, but in the ViC corpus, *wh* adverbs demonstrate greater levels of reducibility, clustering with prepositions and pronouns.

The variances in findings between the analysis here and Altenberg's classes may arise from several factors. The nature of the data used in the two studies differs. Altenberg's results are based on an analysis of a single British English speaker performing a fifty-minute prepared speech passage. The results reported here are based on much more data—albeit limited to monosyllabic words only—from numerous American English speakers in spontaneous, conversational speech. Not only may there be dialectal or other linguistic differences, but also the limitations of Altenberg's dataset size might explain some of the divergent findings. For instance, Altenberg's data only includes nineteen tokens of *wh* adverbs (1987:205), whereas there are 2,419 tokens of *wh* adverbs in the ViC dataset; such a limited set of *wh* adverbs might explain why Altenberg does not find much reduction for this set of words.

Another potential source of the differences between Altenberg's classifications and the clusters presented in §6.2 is that Altenberg makes lexical distinctions that were not included in the hierarchical cluster analysis here. Namely, he separates verb forms of *be*, *have*, and *do* from other "lexical" verbs—e.g., *run*. Although Altenberg provides no explicit reasoning for doing so, the implicit suggestion seems to be that *be*, *have*, and *do* represent a closed class of verbs that might demonstrate differences in stress and reduction when compared to their more lexical, open class counterparts. It is possible that this effect may be due in part to frequency differences between verbs, if *be*, *have*, and *do* are more frequent than other verbs—this possibility of a frequency effect is addressed in §8. It is also possible that the verbs *be*, *have*, and *do* actually are distinct from lexical verbs with respect to their stress properties and patterning, as Altenberg seems to suggest following previous work. In this case, we should see that these verbs are more likely to group together with function word-like clusters than with content word-like clusters.

For the sake of comparison with Altenberg's results, we can split *be*, *have*, and *do* from the other lexical verbs in the cluster analysis. Each of the five verb part-of-speech tags (VB, VBD, VBN, VBP, VBZ) was split into two groups, separating forms of *be*, *have*, and *do*, which I will call the closed class verbs, from other verbs, which I will refer to as the lexical group. Under this reformulation, for example, VBD for past tense verbs became two separate tags: VBD.f for closed class, past tense verb forms of *be*, *have*, and *do* and VBD.c for lexical past tense verbs[49]. For each original verb part-of-speech tag, the closed class verbs represented anywhere from 15 to 62.45% of the data—a breakdown of the data is provided in Appendix C. Third quartile values for standardized phone distance and duration distance measures were recalculated for the separated groups. Results of the reformed hierarchical cluster analysis using the same methodology as described in §6.2 are provided in Figure 22.

---

[49] This, however, does not separate lexical *have* (e.g., *I* have *a dream*) from auxiliary *have* (e.g., *I* have *had a dream*).

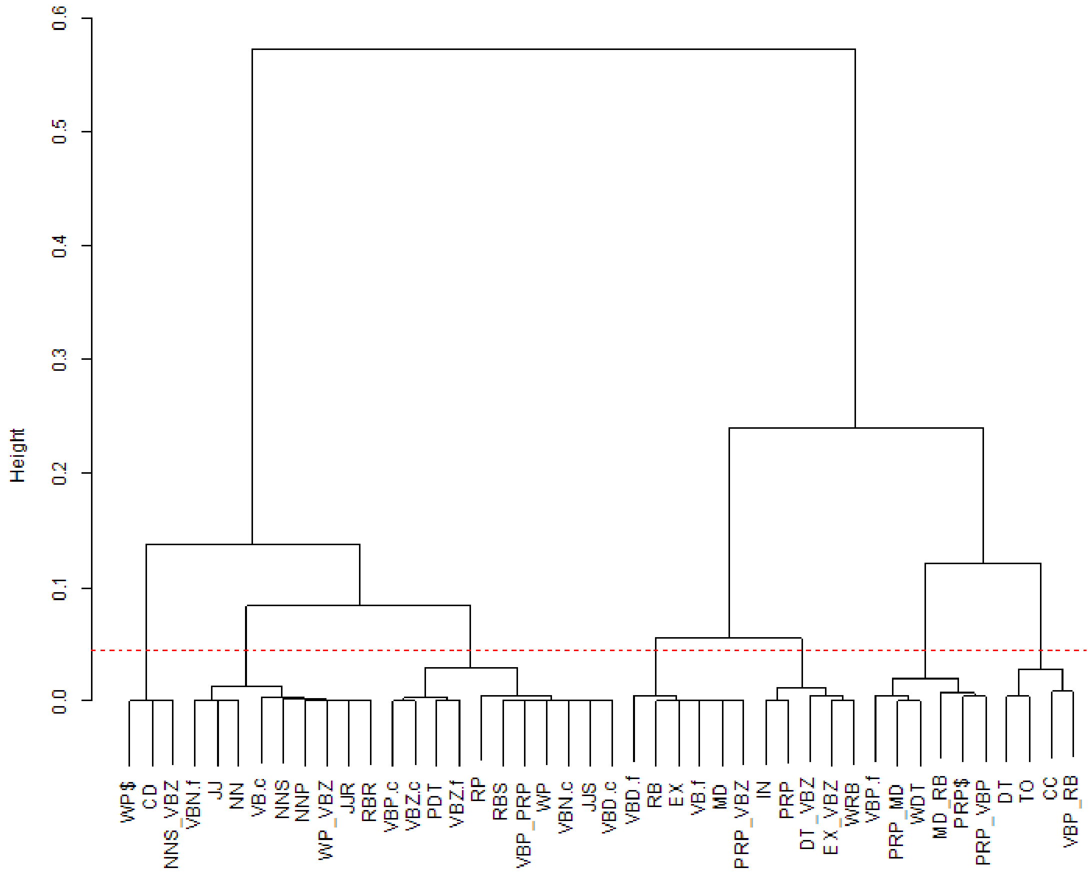

Figure 22. Hierarchical cluster dendrogram: Revised part-of-speech clusters by phone distance and duration distance. Dashed horizontal line represents a seven-cluster solution.

In the revised hierarchical cluster model, there still remains a primary division between a group that contains standardly associated content words (e.g., nouns, verbs, adjectives) in the top left branch and a group that does not contain content words in the top right branch.[50] The low level clusters also remain similar to the original model presented in Figure 20. The main differences between the two hierarchical models are located further up the branching structure. The results in Figure 22 reveal a difference in the two-cluster grouping of content and function words when verbs are divided into closed and open classes. Unlike the first hierarchical cluster model, this model groups

[50] Note that these branches for content and function words are the mirror opposite of those in Figure , due to ordering idiosyncrasies with the `hclust()` and base graphics functions in `R`.

modal auxiliaries and adverbs in the topmost right "function" word branch; recall that in the first cluster model, these part-of-speech categories were located in the main "content" word branch of the dendrogram. Despite being grouped in the function word branch, these words are still closer to the content word category in terms of lack of reduction than other function words, including prepositions, pronouns, and determiners.

The height of the clusters in Figure 22 suggest that a six- or seven-cluster solution is a good place to start in terms of examining word categories with potentially significant differences that are not too granular. Membership in the seven clusters are listed in Table 9. A six cluster solution combines Clusters 4 and 5; a five cluster solution makes no distinction between most of the content words (i.e., combines Clusters 2 and 3). The seven cluster solution is also represented graphically in Figure 23.

Table 9. Seven-cluster hierarchical cluster result groups by part-of-speech and lexical specifications

Cluster	Parts-of-speech
Cluster 1	cardinal number, possessive *wh* pronoun, noun_verb
Cluster 2	adjective, comparative adjective, noun (singular and plural), proper noun, comparative adverb, lexical verb, past participles *done*, *had*, *been*, *wh* pronoun_verb (pres., 3rd ps.)
Cluster 3	superlative adjective, superlative adverb, predeterminer, particle, past tense lexical verb, past participle lexical verb, present lexical verb, *wh* pronoun, present verbs *is*, *has*, *does*, verb_pronoun
Cluster 4	existential *there*, modal auxiliary, adverb, verbs *do*, *have*, past tense verbs *had*, *were*, *was*, *did*, pronoun_verb (pres, 3rd ps.)
Cluster 5	preposition, pronoun, *wh*-adverb, determiner_verb (pres., 3rd ps.), *there*_verb (pres., 3rd ps.)
Cluster 6	possessive pronoun, *wh* determiner, present verbs *are*, *have*, *do*, *am*, modal_adverb, pronoun_modal, pronoun_present verb (not 3rd ps.)
Cluster 7	coordinating conjunction, determiner, preposition *to*, present verb_adverb

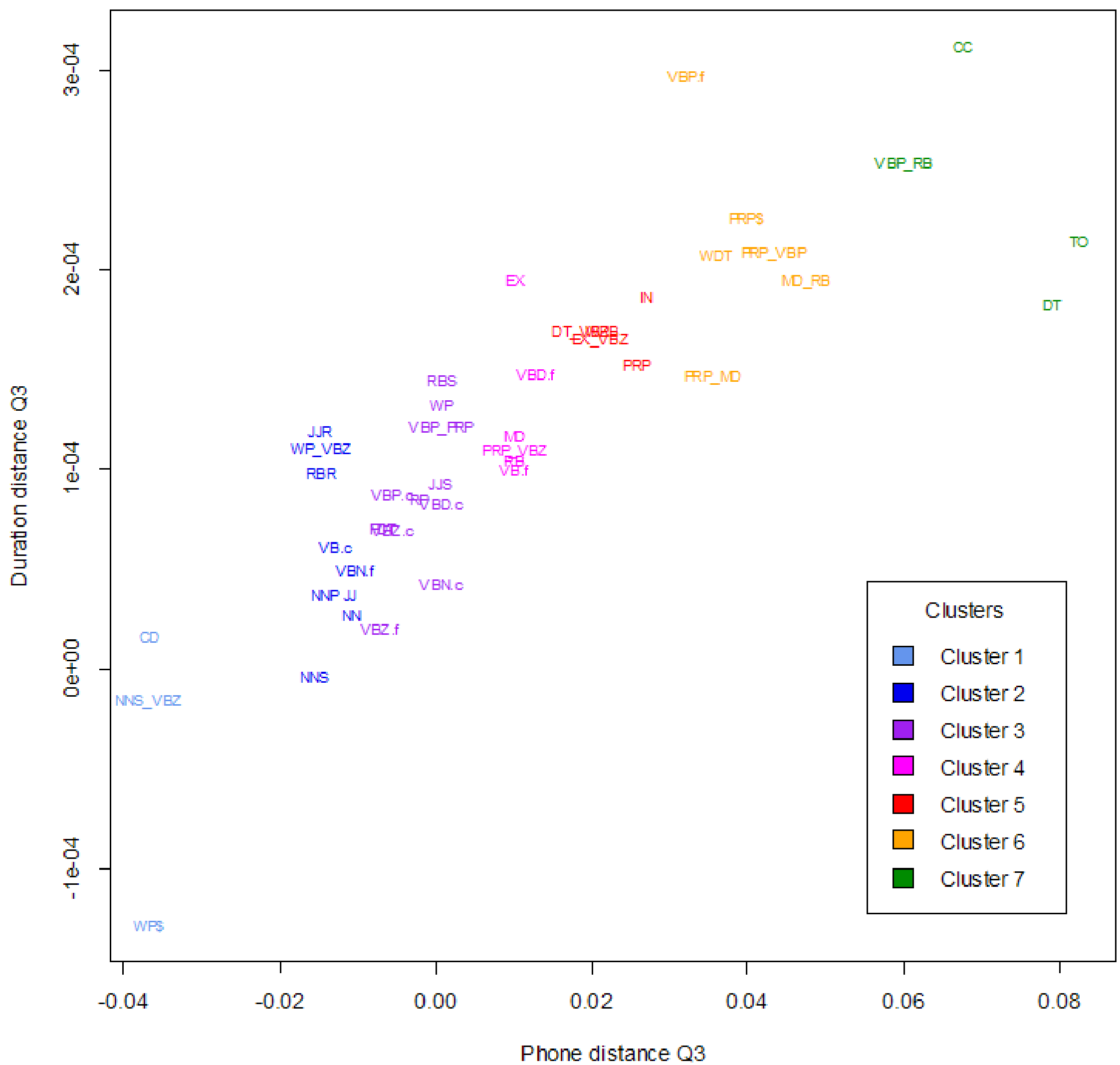

Figure 23. Third quantile (Q_3) values of phone distance and duration distance by hierarchical model part-of-speech and lexically-specific clusters.

Out of the revised hierarchical results, two distinct clusters within the larger function word branch emerge. The group on the left (including Clusters 4 and 5) represent words that do not exhibit as much reduction as the words in the group on the right (Clusters 6 and 7). Such a distinction has similarities to Altenberg's findings, which distinguish between highly reducible function words—for example, determiners, possessive pronouns—and less reducible function words—for example, modal auxiliaries, prepositions.

For verbs, the lexical class of verbs are all limited to the "content" branch of the dendrogram, in Clusters 2 and 3. Closed class verbs (*be*, *have*, *do*), on the other hand, demonstrate a range of reduction patterns: past participles (VBN.f) and third person, present tense verbs (VBZ.f) are clustered with content words while base form verbs (VB.f), present tense, non-third person (VBP.f), and past tense verbs (VBD.f) occur in the "function" branch. It is unclear at this point where the differences in these verb forms come from, whether they are caused by another factor that affects reduction such as phonological structure or frequency (see discussion in §8) or if they represent legitimate differences in the underlying stress properties of these words. Altenberg (1987: 133–134), for his data, notes differences between the closed class verbs, placing forms of *be* lower on the scale of prosodic potential than forms of *have* and *do*. Such an explanation, however, does not appear to be the source of differences in reduction found here, as the verb forms with the least amount of reduction are largely represented by *be* tokens: for example, *is* makes up 77.91% of all closed class, present tense, third person verb tokens, and *been* makes up 74.33% of all closed class, past participle verb tokens. In contrast, present tense, non-third person verbs exhibit the most reducibility, and are made up of a majority of *have* tokens (44.57%).

Given the results from the revised hierarchical model, we can see that some of the differences between the study of reduction in ViC here and Altenberg's results might arise from differences in base groups (i.e., parts-of-speech) and lexical specifications. For example, separating closed class verbs from open class, lexical verbs resulted in the shift of modal auxiliaries from a primarily content word-dominated cluster to a primarily function word-dominated cluster, thus resolving one of the primary differences between the current study and Altenberg's. But, some other divergences in the results, such as the higher prosodic potential of conjunctions in Altenberg, cannot be attributed to such an explanation and might be the result of discrepancies between the types of data or another unconsidered variable between the two studies.

What emerges from the hierarchical cluster models and this comparison with Altenberg (1987) is that there is a scale of reducibility amongst word types, ranging from words that exhibit little propensity for reduction to those that are commonly reduced from their citation forms. The results here, as well as Altenberg's findings, show that words group together along this scale of reducibility, and such a grouping is not strictly binary, as is commonly assumed. Moreover, the results demonstrate that the division between lexical and grammatical words on the basis of syntactic function and of open and closed class distinctions does not directly correlate with a strict division on the basis of phonetic and phonological behavior.

6.4 Interim summary

As stated at the outset of the current section, categories that the hierarchical cluster analysis finds are likely to be meaningful if they

1. improve the power of the analysis of reduction patterns without overfitting (i.e., being too lexically-specific and thus un-generalizable to unseen data);
2. demonstrate independent differences as groups with respect to factors (e.g., frequency, rhythm, structural position) that should distinguish between content and function word classes; and
3. concord with theoretical expectations and/or previous findings.

It has already been discussed in the current section that the cluster analysis produces results similar to those in previous studies based on prosodic potential (e.g., Altenberg 1987). The subsequent sections will address these diagnostics of the clusters further, focusing on the main issue of whether and to what extent the word categories determined by the hierarchical cluster analyses herein are necessary for determining the lexical stress properties of words: that is, are words simply lexically stressed and unstressed based on their content- or function-word-hood, or are the distinctions more finely-tuned?

7 Controls for regression modeling and analysis

The results of the hierarchical clustering analysis in the previous section show that word categories pattern along a scale of reducibility—or "stress potential." Some words—including most content words such as nouns, lexical verbs, and adjectives—have little tendency for segmental and durational reduction as compared to other words—for example, determiners and conjunctions—that reduce more and at higher rates. A general, binary divide between the words was found; however, clustering analysis also uncovered distinct subgroups of words that do not necessarily align with the standard, two-part division of the lexicon into content and function words. Certain standardly-regarded function words such as particles clearly pattern with the groups of words that demonstrate little reduction. Such results are similar to previous conclusions by Altenberg (1987) about the nature of function word prosodic potential. If it is the case that underlying, lexical stress is determined along the lines of lexical categories—stressed content versus stressless function words, as has been largely assumed by the previous literature—then the results presented in the previous section suggest that we need to carefully reconsider these proposed partitions in the lexicon.

The remainder of this chapter addresses the issue of how many clusters—that is, subdivisions of the lexicon—matter for distinguishing word category behavior in terms of lexical stress and reduction. The hierarchical cluster analyses in §5 provide potential solutions with anywhere between two and seven clusters. The questions at hand are to what extent these finer divisions are necessary and meaningful, and to what extent the non-expected behavior (given a traditional binary division) of certain words and parts-of-speech could be otherwise explained by other factors known to affect reduction, resulting in a cleaner, binary lexical division. Based on evidence from linear mixed-effects modeling and information-theoretic model selection procedures, I argue in the following sections that a four-category division of the lexicon—as opposed to the standard binary split or to Altenberg's ten-part scalar divide—is the optimal approach to capturing the differences of reduction patterns between words.

The current section introduces additional factors that are known to affect reduction (§7.1); these factors are combined as part of a base control model in §7.2. Different hypotheses of word categories are compared and discussed in §8.

7.1 Control predictors

Word categories and the presence or absence of lexical stress is by far not the only property known to influence reduction in speech. Other factors that have been shown to affect word pronunciation include but are not limited to speech rate, frequency, predictability, previous experience with a given word, phonological factors, and prosodic or syntactic structure. Thus far, the analysis of word categories presented has not taken into account these known factors of reduction. A question that arises is to what extent membership in the word category clusters uncovered in §6 can be attributed to these other factors, and to what extent word categories of content and function words, in addition to the known factors, are necessary to account for the reduction patterns in the data. Moreover, some of these factors—including frequency, predictability, and structural position—have been previously shown to interact with content and function word categories (e.g., Bell et al. 2009; Selkirk 1996), and will be important in testing whether clusters beyond a standard binary content versus function division are meaningful in §9. The control predictors used in the current study are introduced in turn below.

SPEECH RATE. More reduction should occur at faster speech rates. Speech rate for each word was calculated by taking the average syllables per second of each window of speech from the preceding phrase boundary to the target word and from the target word to the following phrase boundary, excluding the target word itself (following methodology in Frank & Jaeger 2008; a.o.). This measure of speech rate, which has been normalized by each phrase, also helps to control, in part, speaker variations.

FREQUENCY. The connection between reduction and unigram word frequency is well-noted in the literature: more frequent words tend to display reduced pronunciations (Zipf 1929; et seq.).

CONDITIONAL BIGRAM PROBABILITY. Following the idea that the probability of a word is connected to the amount of reduction in pronunciation, previous research has shown that the predictability of a target word given its neighboring words also influences pronunciation (Bell et al. 2003; 2009; a.o.).

Conditional bigram probabilities were calculated as the ratio of the frequency of a word and its previous or following neighbor and the corpus frequency of the neighboring word (i.e., $C(w_i, w_{i\pm1}) / C(w_{i\pm1})$; Jurafsky & Martin (2009:89); a.o.).

ACCENT RATIO. Accent ratio is a measure developed by Nenkova et al. (2007; following Yuan et al. 2005) of how likely a word is to carry pitch accent given how often it has been accented in a training corpus. The hypothesis is that the more often a word has been encountered as accented previously, the more likely it will occur as accented; likewise, the more often a word has been encountered as unaccented previously, the more likely it will occur as unaccented. Though not explicitly suggested by Nenkova et al., the idea of accent ratio—that prior experience with a word will bias accent placement—can be theoretically related to exemplar-based ideas of access and priming.

Accent ratio, as formulated in (52), is the estimated probability that a word carries pitch accent if that probability significantly differs from 0.5 (i.e., chance); if the probability is not significantly different or the word has not been previously encountered, the accent ratio is 0.5. Significance is taken to be $p \leq 0.05$.

(52) Accent ratio of word *w* (from Nenkova et al. 2007:11)

$$\text{Accent ratio }(w) = \begin{cases} \frac{k}{n} \text{ if } \mathrm{B}(k, n, 0.5) \leq 0.05 \\ 0.5 \text{ otherwise} \end{cases},$$

where k = number of instances w has appeared as accented in the training set;
n = total number of instances of w; and
$B(k, n, 0.5)$ = probability under a binomial distribution that there are k successes in n trials if the probability of success and failure is equal.

Trained on a prosodically-tagged portion of sixty Switchboard conversations (Ostendorf et al. 2001), Nenkova et al. report that accent ratio predicted the presence of pitch accent with 75.59% accuracy. Accent ratio alone was found to be a better predictor of pitch accent than any other single predictor that was tested, including phrase position, part-of-speech, bigram probability, unigram frequency, term frequency-inverse document frequency (TF.IDF; i.e., how central a word is in the conversation), and "kontrast" (information structure/focus); moreover, accent ratio was a better predictor than models with several of the other predictors combined (12). Across spoken language corpora, accent ratio was also shown by Nenkova et al. to be a robust predictor of pitch accent: measures of accent ratio trained on the Switchboard corpus predicted pitch accent placement in the Boston University Radio Speech Corpus (Ostendorf et al. 1996) with 82% accuracy (15). While accent ratio is inevitably correlated with other factors of pitch accent placement—for example, many of the low accent ratio words are function words (see Nenkova et al. 2007: Table 3)—, it is worth noting that the accent ratio measure cross-cuts many of these other predictors. For instance, high frequency words can be found to have either high or low accent ratio values, and function words appear in both the high and low accent ratio groups (14).

The accent ratio values used in the current study were taken from Nenkova et al.'s Switchboard-trained set.[51] Words without a pre-assigned accent ratio value by Nenkova et al. were given a value of 0.5. Because pitch accent correlates with stress, we expect to find that accent ratio will correlate with reduction: words with higher accent ratios—indicating words that more frequently occur with pitch accent—should exhibit less reduction and stressability than words with lower accent ratios.

[51] Acknowledgements to Dan Jurafsky for providing the accent ratio measures from Nenkova et al. (2007).

PROSODIC AND SYNTACTIC STRUCTURE. Phrase position—computed either prosodically or syntactically—has been claimed to contribute to reducibility and to distinguish between content and function words. One strong claim in particular is that non-phrase-final elements should undergo more reduction than phrase-final elements, which are more protected from reduction by principles of phrasal stress. Function words, moreover, are expected to be the most susceptible to such phrase-internal reduction effects (Selkirk 1984, 1996; Inkelas and Zec 1993; a.o.; see §2). Phrase position has also been shown in previous corpus studies to be a contributing predictor in some of the best classifier models of pitch accent prominence (Nenkova et al. 2007: 12)

Because there is no syntactic parsing currently available for the ViC corpus, this study controls roughly for phrase-final position. Phrase boundaries were approximated at the ViC silence (<SIL>) and turn-taking (<IVER>) tags, as well as preceding the conjunctions *and*, *but*, *or*, and *if* (following Yao 2011; Anttila 2012). We test here the coarse generalization that increased reduction is expected in non-phrase-final positions, especially for grammatical words. More fine-grained prosodic and syntactic controls are left for future work.

PHONOLOGICAL STRUCTURE. Three controls for phonological (syllabic) structure were used: presence of coda, presence of complex coda, and presence of onset.

WORD. Words themselves may demonstrate individual influences in reduction propensity due to a number of causes (e.g., contextual differences). For example, Bell et al. (2003) reported individual differences in duration and vowel reduction amongst the ten most frequent function words. To control for such effects, random intercepts for orthographic words were included in the mixed-effects models reported here. There are 2,561 unique orthographic words in the dataset.

7.1.1 *Control factors not included*

Not all predictors that are known to affect reduction were included in the analysis, for a variety of reasons. Focus, givenness, and other information content markers (e.g., "kontrast"; see Nenkova et al. 2007) were not controlled for due to limitations in the available semantic coding for the ViC corpus dataset. Phonological neighborhood density (Yao 2011; Gahl et al. 2012; a.o.), the repetition of words (Bell et al. 2003; 2009; a.o.), and other phonotactic conditions (e.g., surrounding phonological environment; Bell et al. 2003; a.o.) have also been shown to affect pronounced durations but were not included here. These predictors are reserved for future investigation, as is any effect on reduction based on speaker variation.

7.2 Control models

Two control models are introduced here, with phone distance and duration distance as the dependent variables. The models here should demonstrate to what extent the controls, as introduced in §7.2, predict reduction and stressability in the ViC dataset of monosyllabic words. The potential added contribution of content-function word categories for explaining reduction patterns follows in §8.

Generalized linear mixed-effects models were fit using the `glmer` function from the `lme4` package in `R` (Bates et al. 2013). All variables except for multi-level, categorical predictors were centered. Numerical variables were also standardized by dividing by twice their standard deviations (following Gelman 2008; a.o.). The numerical variables of speech rate, word frequency, conditional bigram probabilities, and duration distance[52] were log-transformed prior to standardization. The control model for phone distance is presented in §7.2.1, and the control model for duration distance is presented in §7.2.2.

---

[52] Because log(0) is undefined, duration distance (range = [-0.3000327, 0.1084973]) was first shifted by +0.3000328 before log transformation.

7.2.1 *Control model: Phone distance*

The results of the control model with phone distance as the dependent variable are given in Table 10. Partial effects of the fixed predictors are plotted in Figure 24 by holding all other numerical predictors at their medians and categorical predictors at their reference levels.

Table 10. Control model for phone distance (mixed-effects generalized linear model)

Factor	Estimate	Std. Error	t value
Intercept	-0.005277	0.0009636	-5.476
log(Speech rate)	0.002655	0.0003911	6.788
log(Word frequency)	0.003986	0.0002392	16.666
log(Bigram previous)	0.001289	0.0000546	23.612
log(Bigram next)	0.001317	0.0000501	26.308
Accent ratio	-0.1856	0.02652	-6.997
Phrase position = nonfinal	0.005533	0.0002228	24.838
Coda = Y	0.01565	0.001138	13.749
Complex coda = Y	0.002589	0.0003784	6.843
Onset = Y	-0.007838	0.001466	-5.348
Interactions			
log(Freq) * log(Bigram prev)	0.0003419	0.0000242	14.102
log(Freq) * log(Bigram next)	0.0004595	0.00002024	22.709
Groups	Name	Variance	Std. Deviation
Word	(Intercept)	0.00015481	0.12442
Residual		0.00100794	0.031748
N	206858	Groups: word	2561
Log likelihood	418847	R^2	0.4279625
Deviance	-837870	AIC_c	-837666.1

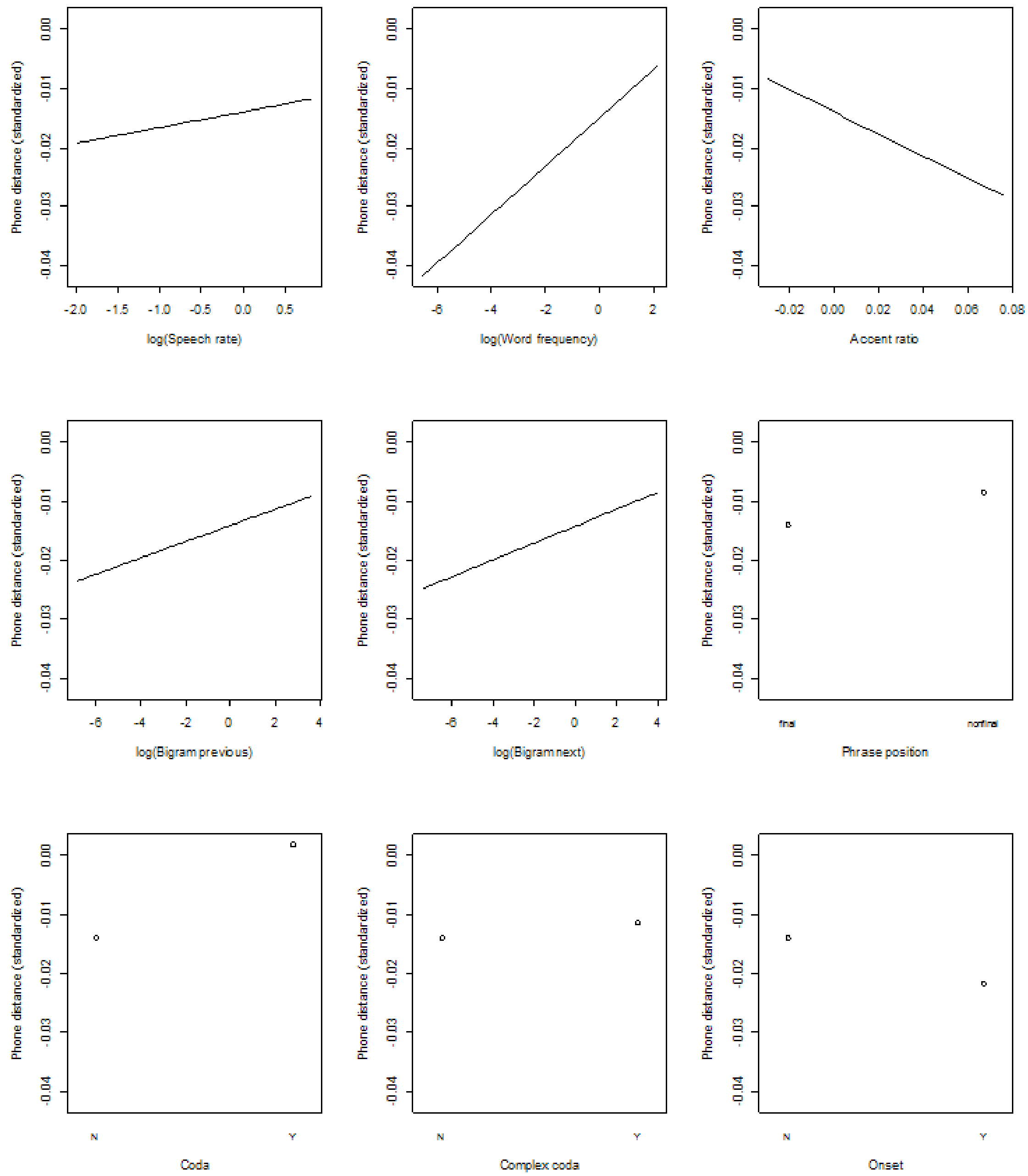

Figure 24. Partial effects of control model for phone distance (see Table 10).

All control predictors, with the exception of onset, behave as expected. As speech rate, unigram frequency, and bigram predictability (both previous and following) increase, so does the amount of segmental reduction, as the positive regression coefficients demonstrate. Of these four predictors, word frequency[53] has the greatest magnitude of effect: a one percent increase in word frequency indicates about a 0.4% increase in phone distance (±0.02%). As expected, accent ratio is inversely correlated with phone distance (β = -0.1856): greater accent ratio values predict less segmental reduction (i.e., lower phone distance values).

For structural factors, the presence of a coda or a complex coda correlates with more segmental reduction. These effects are as expected, under the hypothesis that coda segments—in contrast to nuclear vowels and onset segments—are the most easily reduced or deleted. We also might expect that the presence of an onset would lead to more reduction; however, the model results suggest that the presence of an onset actually correlates with less reduction when measured by phone distance. One possible explanation for such an effect is that onsets tend to remain segmentally faithful—perhaps for structural well-formedness or informativity purposes—but can reduce in other ways: for instance, the control model of duration distance in §7.2.2 will show that the presence of an onset still correlates with more durational reduction in monosyllabic words. Finally for structural factors, there is also a simple main effect of phrase position in the expected direction. Words that occur in non-phrase-final position tend to exhibit more segmental reduction as compared to words in phrase-final positions. This concurs with the theoretical hypothesis that reduction is in part determined by prosodic phrasing, with phrase-final position being privileged for more stress assignment due to phrasal prominence (e.g., Selkirk 1996).

---

[53] Word frequency is not exactly a linear predictor. Using a five-knot restricted cubic spline for frequency improves the model significantly. For interpretability, however, and because we are not primarily interested in word frequency here, I assume a log-linear relationship for the time being.

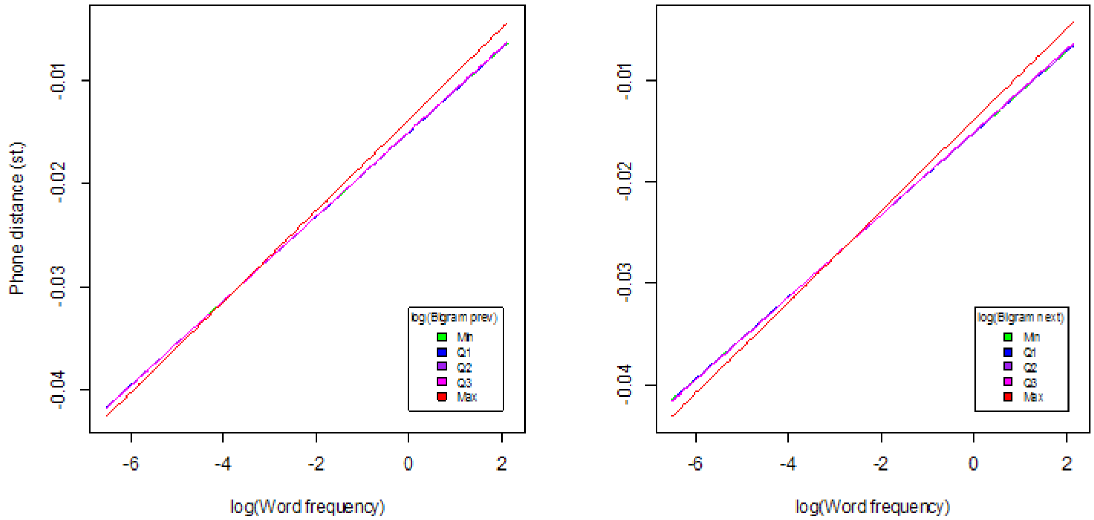

Figure 25. Interactions of control model for phone distance (see Table 10). (Left to right) (a) log(Word frequency) * log(Bigram prev); (b) log(Word frequency) * log(Bigram next). Lines represent minimum, maximum, and quartiles of the numerical predictors.

Because Bell et al. (2009) found a significant interaction between unigram frequency and *n*-gram predictability, interaction terms between frequency and conditional bigram probabilities were included in the control models. The interactions of the model are plotted in Figure 25 and show that the effects of frequency and bigram probabilities are compounded. As the values for bigram probabilities increase, as illustrated in the plots at the quartiles, they pull up the effect of frequency effect, indicating that there is more reduction when the word is both frequent and highly predictable given its immediately adjacent words.

7.2.2 Control model: Duration distance

The results of the control model with duration distance as the dependent variable are given in Table 11. Partial effects of the fixed predictors are plotted in Figure 26 by holding all other numerical predictors at their medians and categorical predictors at their reference levels.

Table 11. Control model for duration distance (mixed-effects generalized linear model)

Factor	Estimate	Std. Error	t value
Intercept	0.0002589	0.00009473	2.73
log(Speech rate)	0.002019	0.00003882	52.01
log(Word frequency)	0.0003892	0.0000236	16.49
log(Bigram previous)	0.0001947	0.00000542	35.93
log(Bigram next)	0.0002628	0.00000497	52.90
Accent ratio	-0.008709	0.002608	-3.34
Phrase position = nonfinal	0.003021	0.00002211	136.64
Coda = Y	0.0007012	0.0001123	6.24
Complex coda = Y	0.0001366	0.00003751	3.64
Onset = Y	0.0007317	0.0001444	5.07
Interactions			
log(Freq) * log(Bigram prev)	0.0000332	0.00000241	13.81
log(Freq) * log(Bigram next)	0.00006665	0.000000201	33.20
Groups	Name	Variance	Std. Deviation
Word	(Intercept)	0.0000014936	0.0012221
Residual		0.0000099322	0.0031515
N	206858	Groups: word	2561
Log likelihood	896660	R^2	0.2776217
Deviance	-1793552	AIC_c	-1793293

The results of the control model for duration distance are similar to those for phone distance. Shorter durations correlate with increased speech rates, word frequency, and predictabilities. Increase in accent ratio—probability that a word carries pitch accent in the Switchboard training dataset (Nenkova et al. 2007)—indicates less reduction and longer durations, while lower accent ratio values indicate more reduction. All structural predictors behave as expected. The presence of codas, complex codas, and onsets all contribute to shorter word durations. Unlike the phone distance control model (§7.2.1), having an onset increases—rather than decreases—the distance from estimated citation durations to the actual pronounced durations by 0.07%, suggesting shorter pronunciations than expected compared to words without onsets.

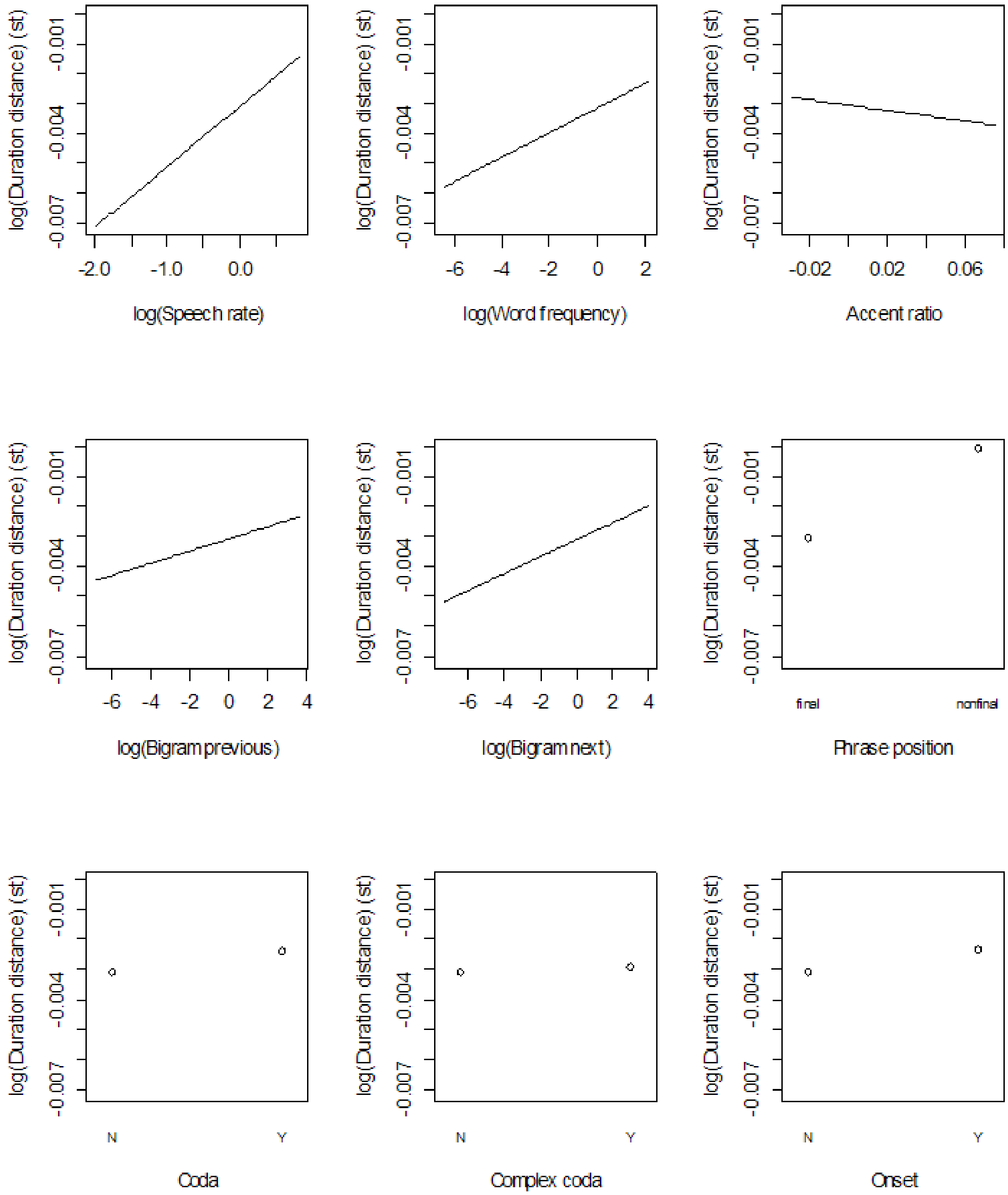

Figure 26. Partial effects of control model for duration distance (see Table 11).

As with the control model for phone distance, interactions between bigram probabilities and frequency were included and show that the combination of increasing bigram probability and increasing frequency leads to even more duration reduction than the simple effects alone. Because the results are the same as with phone distance, I do not show plots of the interactions here.

8 Testing content-function word categorizations

With the control models presented in §7, I now return to the issue of lexical versus grammatical word categories. The issues at hand are [1] what divisions of the lexicon and how many are necessary based on differences in reduction, and [2] whether these divisions are additionally motivated by known differences between word categories. In addressing the former issue, there are multiple hypotheses for word clusters to consider from the results of the cluster analyses in §6, in addition to the standard content versus function split and a highly specific ten-point scale, as proposed by Altenberg (1987). These hypotheses are tested in this section using mixed-effects regression modeling and information-theoretic model comparison, controlling for other known predictors of phonetic reduction. The methodological details are introduced in §§8.1–8.2 and the results in §8.3. Based on the resulting model, the latter issue of word category interactions is taken up in §9. A discussion of the results follows in §9.3.

8.1 Word categorization hypotheses

We start at the outset of this section with multiple hypotheses available about the divisions between content and function words in the lexicon. Divisions from the previous literature include the standard binary content versus function word categorization as well as more fine-grained proposals such as Altenberg's (1987) ten-point scale based on prosodic potential. In §6, the hierarchical cluster analyses provide sets of potential word clusters based on amounts of observed reduction by part-of-speech. These clusters range from binary classifications in 2-cluster solutions to possible solutions containing as many as seven meaningful clusters (e.g., Table 9). There is also the null hypothesis that must be considered: that no word classification is necessary, and reduction in speech can alone be predicted by the control factors discussed in §7.1.

Thirteen competing hypotheses of content-function word categorization are tested here, drawn from the two hierarchical cluster models presented above and from the previous literature. These hypotheses are listed in (53), with a full list of cluster membership in Appendix D. For ease of reference, I call the first hierarchical cluster model from §6.2, which makes no distinction between lexical and closed-class verbs, as H-Clust A, and the second hierarchical cluster model (§6.3), which distinguishes between lexical and closed-class verbs, as H-Clust B. No 3-cluster solution is tested for H-Clust A because there is no clear 3-cluster solution cut of the hierarchical dendrogram (see Figure 20).

(53) a. H-Clust A: 2 cluster, 4 cluster, 5 cluster
b. H-Clust B: 2 cluster, 3 cluster, 4 cluster, 5 cluster, 6 cluster, 7 cluster
c. Standard content vs. function binary divide
d. 10-part scale (Altenberg 1987)
e. Hybrid: content vs. 5 cluster function
f. Null control model

It is assumed here that word categorization falls along a sequential scale of reducibility, following the common assumption that content words have lexical stress and are less susceptible to reduction. To test the differences between each category, reverse Helmert coding (i.e., difference contrast coding) for the factor levels was used (`contr.helmert()`). Reverse Helmert coding compares each factor level to the mean of previous levels, which is appropriate for sequential factor levels as the word clusters tested here are taken to be.

8.2 Information-theoretic model selection

For each clustering hypothesis listed in (53), two models—one predicting phone distance and the other predicting duration distance—are tested based on the control models presented in §7.2. The resulting thirteen models for each dependent variable are compared using an information-theoretic model selection method based on Akaike Information Criteria (AIC) (Anderson & Burnham 2002; Burnham & Anderson 2002; 2004; following Akaike 1973, et seq.). This section briefly introduces the methodology; results and discussion follow in §8.3.

AIC-based model selection is founded on the idea that all models are mere approximations of full reality, an ideal for which the parameters (θ) are unknown. Because we are only estimating full reality, every model experiences some amount of information loss when compared to the "true model" of reality: there will always be an amount of uncertainty that does not capture full truth. AIC measures are crucially related to the metric of Kullback-Leibler information distance between full reality and the estimated model (Kullback & Leibler 1951), and they measure how much information is lost. Comparing the differences between AICs of candidate estimated models provides a way to quantify the relative strength of evidence in favor of each candidate model. In this way, we can compare how well each clustering hypothesis produces parsimonious models that suffer the least amount of information loss while using the fewest number of parameters.

AIC is calculated based on model maximum log likelihood. Here, second order AIC, AIC_c, is used, which adjusts for the number of parameters in a model.[54] The formula for AIC_c is given in (54).

(54) $\mathrm{AIC_c} = -2\log\left(\mathcal{L}\left(\hat{\theta}\middle|data\right)\right) + 2K,$
where $\mathcal{L}\left(\hat{\theta}\middle|data\right)$ is the likelihood of the observed data given parameters $\hat{\theta}$, and K is number of estimable parameters in the model.

Algebraically lower AIC_c values indicate less information loss because they are inversely correlated with increasing model likelihoods. For the current study, AIC_c was

[54] Second order AIC_c is intended for small n model comparison and dovetails with the standard AIC measure as sample size increase; therefore, it is more conservative to use AIC_c for model comparison in general, regardless of sample size (Burnham and Anderson 2004: 269–270).

calculated using the `AICc()` function of `R` package `MuMIn` (Bartoń 2013). There is some discussion of the appropriateness of AIC comparison for linear mixed-effects models because the estimation of degrees of freedom for models with random effects is somewhat unclear; however, because we are interested only in the differences between fixed effects here, the use of AIC_c should pose little issue (cf. conditional AIC_c, which is recommended when the comparison of random effects is of interest; see discussions in Vaida and Blanchard (2005); Greven and Kneib (2010); a.o.).

Unlike other methods of model selection (e.g., anovas, *p*-values, R^2), the use of comparative AIC_c does not assume nested models (Burnham & Anderson 2004:267): only the dependent variables and data must remain the same across compared models. This feature of AIC_c model selection makes it particularly useful in the case undertaken here, where the tested hypotheses of word categorizations are closely related but, crucially, are not nested measures (e.g., the ten-point Altenberg scale does not use the same clusters as a four-cluster solution produced by the hierarchical cluster analysis in the above sections). For similar applications within linguistics of model selection and its related method of model averaging (not used here), see Grafmiller & Shih (2011) and Kuperman & Bresnan (2012).

A methodological point to address is why one does not take the approach of a "dummy," brute force, exhaustive search of the entire space of possible clustering options—rather than limiting ourselves to thirteen tested hypotheses—to find the model and associated cluster solution that produce the lowest AIC_c value. To do so would be an abusive misuse of information-theoretic model selection strategies, which Anderson and Burnham strongly warn against (2002; see also Burnham & Anderson 2004). Instead, model selection should be guided in part by human logic and well-supported theoretical motivations. The hierarchical cluster analyses help serve the purpose of narrowing the search space by suggesting cluster solutions that we have seen to be potentially meaningful—that is, they align (for the most part) with our expectations of function-content reduction behavior (see discussion in §6). A brute force search of all possible partitions of the thirty-nine parts-of-speech in the dataset, which amounts to upwards of 10 decillion possible cluster solutions,[55] would be untenable not only for theoretical reasons but also on the grounds of computational limitations and statistical power.

8.3 AIC$_c$ results and discussion

Before presenting the AIC$_c$ model comparison results, let us first briefly examine a model of phone distance that utilizes a standard binary content versus function word classification, in addition to the control predictors (Table 12):

Table 12. Content versus function word model for phone distance

Factor	Estimate	Std. Error	t value
Intercept	-0.007698	0.001061	-7.254
log(Speech rate)	0.002591	0.0003911	6.625
log(Word frequency)	0.003488	0.0002487	14.023
log(Bigram previous)	0.0008959	0.0001049	8.538
log(Bigram next)	0.001164	0.0009949	11.696
Accent ratio	-0.174	0.02682	-6.488
Phrase position = nonfinal	0.003556	0.0003418	10.406
Coda = Y	0.0155	0.001131	13.698
Complex coda = Y	0.002623	0.000378	6.939
Onset = Y	-0.007153	0.00146	-4.9
Word cluster = function	0.002799	0.0007894	3.545
Interactions			
log(Freq) * log(Bigram prev)	0.0002551	0.00003191	7.995
log(Freq) * log(Bigram next)	0.000382	0.00002773	13.775
log(Freq) * Cluster = fx	-0.00143	0.0005273	-2.711
log(Bigram prev) * Cluster = fx	0.003316	0.0004447	7.456
log(Bigram next) * Cluster = fx	0.0001847	0.0001257	1.470
Phr Pos = nonfin * Cluster = fx	0.0005731	0.0001321	4.338
Groups	Name	Variance	Std. Deviation
Word	(Intercept)	0.00015098	0.012288

[55] As calculated using Bell numbers (partitions for a set), the thirty-ninth Bell number for thirty-nine parts of speech would amount to 10,738,823,330,774,692,832,768,857,986,425,209. Hat tip to Richard Conway for calculating this number via Java script.

Residual		0.00100756	0.031742
N	206858	Groups: word	2561
Log likelihood	418865	R^2	0.4281577
Deviance	-837978	AIC_c	-837692.9

What is crucial to note here is that the addition of a simple content versus function word classification to the control model (§7.2.1: Table 10) significantly improves its explanatory accuracy ($\chi^2 = 108.65$, df = 5, $p < 0.001$). Note that there is also a large difference of 26.8 between the AIC_cs of both models[56], in favor of the model that includes a content-function categorization of words. This difference confirms the assumption that word categories—in addition to the control variables are necessary in predicting stress and reduction. The word categories also demonstrate interactions with other factors of reduction such as frequency, predictability, and structural position; I will return to the discussion of these interactions in §9.

AIC_c results for the tested models are given in Table 13 (phone distance models) and Table 14 (duration distance models). More negative AIC_c values indicate less information loss. Models are ranked in the tables in order of increasing AIC_c—that is, best to worst, given the set of models and the empirical data. Δ_i indicates the difference between AICs of the best model out of the candidate set (i.e., candidate with the minimum with AIC) and the given model *i*. Note that the AIC_c values reported here are negative, resulting from large positive log likelihoods: this is due to linear regression and low dispersion in the data. Typical log likelihoods are negative (with positive AIC_c) when estimating categorical predictors in logistic regression (because the range of response is between 0 and 1). *k* is the number of fixed parameters in the model.

[56] As a rule of thumb, AIC_c differences > 10 are considered large (Burnham and Anderson 2004:271), though we will primarily be interested in model rankings here rather than concrete difference (Δ_i) values.

Table 13. Second order Akaike Information Criteria and weights for phone distance

Cluster model	Log lik	k	AIC_c	Δ_i
4 clusters: hclust A	418991	27	-837924.3	0
5 clusters: hclust B	418971	32	-837873.2	51.1
3 clusters: hclust B	418951	22	-837854.7	69.6
5 clusters: hclust A	418961	32	-837853.2	71.1
2 clusters: hclust B	418924	17	-837810.5	113.8
2 clusters: hclust A	418912	17	-837785.5	138.8
4 clusters: hclust B	418921	27	-837784.7	139.6
6 clusters: hclust B	418930	37	-837782.6	141.7
Hybrid model	418926	37	-837774.5	149.8
7 clusters: hclust B	418904	42	-837719.9	204.4
2 clusters: content/function	418865	17	-837692.9	231.4
Null model	418847	12	-837666.1	258.2
10 clusters: Altenberg	418411	152	-836714.6	1209.7

Table 14. Second order Akaike Information Criteria and weights for duration distance

Cluster model	Log lik	k	AIC_c	Δ_i
2 clusters: hclust B	896709	17	-1793381	0
2 clusters: content/function	896698	17	-1793358	23
2 clusters: hclust A	896683	17	-1793327	54
3 clusters: hclust B	896686	22	-1793325	56
4 clusters: hclust A	896680	27	-1793301	80
Null model	896660	12	-1793293	88
5 clusters: hclust B	896651	32	-1793235	146
4 clusters: hclust B	896642	27	-1793227	154
5 clusters: hclust A	896636	32	-1793204	177
Hybrid model	896635	37	-1793192	189
6 clusters: hclust B	896630	37	-1793182	199
7 clusters: hclust B	896604	42	-1793121	260
10 clusters: Altenberg	869302	52	-1792495	886

In general, the results of the AIC_c comparisons show that clustering hypotheses with more clusters do not do as well as hypotheses with fewer clusters. For instance, the seven-cluster solution from H-Clust B does not improve the model likelihoods of predicting phone distance or duration distance. Likewise, the six-cluster solution as well as the hybrid hypothesis, which marries a five-cluster solution with a general content-versus-function word split, did not do as well as models with fewer clusters across both model rankings. The lower rankings of high-numbered cluster solutions in the AIC_c results suggest that finer-grained distinctions within content and function words may be unnecessary in modeling the reduction facts, and that some of the finer distinctions of reduction that we saw in the hierarchical cluster models can be explained by the control predictors that are included here. In other words, too many clusters leads to over-specification of word categories that are not warranted given the available evidence.

As we saw above, the null, control model is improved by the addition of word categories. Both model rankings demonstrate multiple hypotheses that have smaller AIC_c values when compared to the control model. In the phone distance model ranking (Table 13), the null model ranks below all of the word categorization solutions, with the exception of Altenberg's ten-part scale. For duration distance models, the null model proves to better than several models with more divisions in word categories; however, it still does not rank as highly as the models with fewer divisions in word categories.

Despite the similarities we noted between Altenberg's scale and the cluster analysis results in §6, the ten-part scale as proposed by Altenberg receives little evidentiary support when compared to the other models considered. As already mentioned, divisions with fewer clusters are generally preferred for predicting reduction, and Altenberg's scale falls on the extreme end of having many clusters (see Table 4 and Appendix D): it seems as though having so many divisions is over-specified given the control predictors and the distribution of the data. Another reason that Altenberg's scale did not produce a good fit for the data might be that because he gave no indication of where contractions should fall on the spectrum of prosodic potential, the contractions in the ViC dataset (n = 13,503) were lumped into their own category here. Were the contractions more clearly defined and divided into the ten-part scale, it is possible that the model's performance would improve; however, we do not expect it to improve to the same level as the models with low-numbered cluster solutions, because a ten-part division is too overly specific for the data on hand.

The two model rankings for phone distance and duration distance models differ primarily in their treatment of models with fewer clusters in word categorization. For phone distance models, clusters in the middle range of number of word categories (e.g., four-cluster solutions) had lower AIC_c values than models with clusters at either end of the range (e.g., seven-cluster solutions and two-cluster solutions). On the other hand, duration distance models preferred fewer numbers of cluster overall: in Table 14, two- to four-cluster solutions tend to perform better than others with more specific divisions in the lexicon.

The difference between these two rankings can stem from multiple factors, including the distribution of the data and the nature of the type of reduction that is measured. Duration distance measures extremely small, normalized differences between estimated citation durations and actual pronounced durations, resulting in the majority of the data being located within a narrow range (i.e., low dispersion) of the scale. Consequently, a coarse distinction in word categories will expectedly produce more reliable results because coarse distinctions can more easily pick out extreme differences between words that are more noticeable in under-dispersed data.

The segmental reduction model ranking differs perhaps because the data is not as narrowly confined on the scale as duration distance values. Moreover, segmental reduction is a more drastic form of reduction than duration differences because of the categorical faithfulness violation incurred when deviating from one underlying phoneme to a different surface phone. Duration distance does not necessarily indicate such categorical phonetic changes: for example, a difference in milliseconds between two pronunciations of the same phone [s] is likely a smaller difference than meaning to pronounce the phone [s] and actually pronouncing [z], [f], or [θ]. Because the differences are not as minute, it is potentially easier to distinguish between the differences, and segmental reduction measures produce clearer tests of cluster solutions that are not as biased towards finding extreme solutions as the duration distance model is.

Taking into account the differences between both models, the rankings suggest that a four-cluster solution from H-Clust A (henceforth four-cluster A), shown in Table 15 below, is the most promising division of word categories.

Table 15. Four-cluster hierarchical cluster result groups by part-of-speech (H-Clust A)

Reducibility	Cluster	Parts-of-speech
Less reducibility	Cluster 1	cardinal number, possessive *wh* pronoun, adjective, comparative adjective, noun (singular and plural), proper noun, predeterminer, comparative adverb, verb, present verb (3rd ps.), *wh* pronoun_verb (pres., 3rd ps.), noun_verb
↑	Cluster 2	existential *there*, superlative adjective, modal auxiliary, adverb, superlative adverb, particle, past participle verb, past tense verb, present verb (not 3rd ps.), *wh* pronoun, pronoun_verb (pres., 3rd ps.), verb_pronoun
↓	Cluster 3	preposition, pronoun, pronoun (possessive), *wh* determiner, *wh* adverb, determiner_verb (pres.), *there*_verb (pres.), modal_adverb, pronoun_modal, pronoun_verb (pres., not 3rd ps.)
More reducibility	Cluster 4	coordinating conjunction, determiner, preposition *to*, verb_adverb

For phone distance models (Table 13), the four-cluster A solution ranks the highest out of all the models tested; the difference in strength of evidence (Δ_i) between this model and the next highest-ranking model is 51.1. For duration distance models (Table 14), the four-cluster A solution is the highest ranking amongst mid-range cluster solutions. The model with the four-cluster A is the only model to rank above the null control model for duration distance, with the low-numbered clusters (e.g., two-cluster A) that do well for duration distance.

For both model rankings, a model with four-cluster A does better than a model with four-cluster B because the later makes more extreme divisions in the content word branch of the hierarchical cluster model: for instance, the first cluster in the four-cluster B solution (see Appendix D) consists solely of cardinal numbers, possessive *wh* pronouns, and noun_verb compounds. In contrast, the first cluster in the four-cluster A solution more evenly includes prototypical content-like words (i.e., nouns, verbs, adjectives). As discussed above, it is unlikely that cardinal numbers, possessive *wh* pronouns, and noun_verb compounds form any special subclass distinct from protoypical content words because two of these parts-of-speech have very low frequency; additionally, the difference between cardinal numbers and content words could be due to other factors of reduction, as accounted for by the control predictors.

The other two main differences between the four-cluster A and four-cluster B solutions is the division of the function word branch of the hierarchical cluster analysis. In H-Clust A, prepositions, pronouns, and possessive pronouns form one category, separated from modal auxiliaries, particles, and more content-like material (e.g., past-tense verbs). H-Clust B instead classifies possessive pronouns with the lowest cluster of determiners and conjunctions that should exhibit the most reduction. In H-Clust B, modal auxiliaries, adverbs, and past tense verbs are all considered part of the function word branch, which does not accord with our expectations of function and content word divisions based on other semantic and syntactic factors.

The second difference in function word categorization between the four-cluster A and B models is that in H-Clust A, coordinating conjunctions, determiners, the preposition *to*, and verb_adverb compounds form a separate cluster, distinct from the other function words (e.g., prepositions, pronouns, etc.). H-Clust B, on the other hand, groups all of these function words into one large cluster at the bottom of the reduction scale. The separation of conjunctions and determiners into their own category, as in H-Clust A, falls in line with previous findings (e.g., Altenberg 1987). In particular, determiners are known to be an extreme class of reducers, and giving these words their own prosodic category makes more sense than giving, for example, cardinal numbers their own prosodic category at the top of the reduction scale.

The differences in function word categorization in particular turn out to be crucial. If we test models using a revised H-Clust B solution that makes a more balanced distinction at the top of the scale (i.e., combines cardinal numbers, possessive *wh* pronouns, and noun_verb compounds with the other content words), then the revised B models, which only differ from the A models in the categorization of function words, still exhibit significantly lower AIC_c values. A phone distance model with the revised H-Clust B solution has an AIC_c of -837767.1, ranking it lower than the hybrid model and both of the original four-cluster solutions in Table 13. A duration distance model with the revised H-Clust B solution has an AIC_c of -1793278: this AIC value indicates that the model performs better than the original H-Clust B hypothesis but still does worse than even the null, control model with no word categorizations (see Table 14).

The overall results of the AIC_c rankings suggest that, based on patterns of reduction, an optimal word categorization is one that is more finely-tuned than a binary content versus function word division but not so fine-tuned as to be too overly specific with unnecessary divisions in the lexicon. A systematic search of representative cluster hypotheses shows a four-cluster solution from the hierarchical cluster analysis presented in §6.2 (H-Clust A; see also Table 15) to be one of the highest ranking options that were tested across both segmental and durational reduction measures. Theoretically, the categories as proposed by this cluster solution make sense: there is a larger division (between clusters 2 and 3) that accords with a traditional content versus function word division, and each branch features another split that separates the more prototypical members from those that fall more towards the middle of the reducibility scale. Known reducers, such as determiners, are given their own prosodic category in cluster 4 while highly prototypical content words including nouns, base form verbs, present tense verbs, and adjectives are given a prosodic category representing words that reduce far less than the other categories. Having more than four clusters results in over-specified models that do not perform as well, which suggests that the more fine-grained differences observed in §6 and in previous work (e.g., Altenberg 1987; Nenkova et al. 2007) might be accounted for by other factors known to affect reduction. Once these controls are included, four word categories seems to be what the evidence shows is necessary.

While these categories make theoretically intuitive sense, a critical test for whether all four word categories are meaningful is to look at whether they behave independently in terms of interactions with factors—for instance, frequency, predictability (Bell et al. 2009)—that are known to be sensitive to content and function word differences. This test is taken up in §9, and we will also test the independence in behavior of these categories with respect to rhythmic environment in the next chapter.

Before moving on to interactions, there is one more unresolved issue to settle, which concerns the clustering of content and function words in shared categories. Given that the majority of previous literature has upheld a strict division between content and function words, to what extent is this division justified based on reduction patterns in spoken language? The AIC_c rankings demonstrate that mixed-class two-cluster solutions, which combine certain function words within the content word cluster, perform better than a more standard content versus function word division (with the sole exception of the two-cluster A model with respect to duration distance). We can also examine the results of the model for phone distance that tests a hybrid classification scheme, with a distinction between content words and function words and a scalar division within function words. The relevant predictors from the model are given in Table 16[57] below; a full model is provided in Appendix E. Duration distance model results are comparable and, for brevity, are not provided here.

What the hybrid models reveal is that there is little significant difference between the reduction patterns of content words (reference level at the intercept) and the first two clusters of function words, which are highlighted in grey in Table 16 and contain predeterminers, *wh* pronouns, and particles. There are a couple minor trends towards significance in terms of interactions with frequency (log(Freq) * Cluster = Fx 2, p_{MCMC} = 0.0452) and bigram probability give the following word (log(Bigram next) * Cluster = Fx 1 , p_{MCMC} = 0.0394), but these differences are much less than the ones exhibited by more prototypically function word clusters, such as Fx 4 or Fx 5, which contain determiners, pronouns, and prepositions. Crucially, the control predictor of phrase position does not single out function words in the first two clusters as behaving any different from their content word counterparts. We take this result as evidence that certain function words, like particles, should not be classed separately from content words that exhibit the same behavior in reduction. The more extreme differences in reduction patterns are seen only with the more prototypical function words (e.g.,

---

[57] *P*-values (MCMC and *t*) obtained via Markov chain Monte Carlo (MCMC) sampling, using `pvals.fnc()` from `R` package `languageR` (Baayen 2013).

determiners, etc.), and only then do the differences call for separation of lexical categories that are primarily made up of function words.

Table 16. Hybrid cluster regression model for phone distance (condensed: not all predictors shown)

Factor	Est	Std. Err	t value	pmcmc	Pr (>|t|)
Intercept	-0.008	0.00123	-6.451	0.0001	0.0000
log(Word frequency)	0.0014	0.00061	2.263	0.0128	0.0236
log(Bigram previous)	0.0013	0.00009	14.506	0.0001	0.0000
log(Bigram next)	0.001	0.00009	11.514	0.0001	0.0000
Phrase position = nonfinal	0.0056	0.00044	12.598	0.0001	0.0000
Word cluster = Fx 1	-0.003	0.00228	-1.207	0.1758	0.2275
= Fx 2	-0.0007	0.00087	-0.853	0.5034	0.3938
= Fx 3	0.0021	0.00042	4.907	0.0001	0.0000
= Fx 4	0.0018	0.0003	5.953	0.0001	0.0000
= Fx 5	-0.0002	0.00026	-0.907	0.4744	0.3642
…					
Interactions					
log(Freq) * Cluster = Fx 1	0.0001	0.0012	0.118	0.9222	0.9058
= Fx 2	-0.0013	0.00062	-2.133	0.0452	0.0329
= Fx 3	-0.00008	0.00027	-0.28	0.9236	0.7794
= Fx 4	0.00006	0.00019	0.312	0.839	0.7547
= Fx 5	-0.0007	0.00018	-3.92	0.0001	0.0001
log(Bigr prev) * Cluster=Fx1	-0.00004	0.0002	-0.214	0.8435	0.8305
= Fx 2	0.00001	0.0001	0.020	0.9888	0.9842
= Fx 3	0.0001	0.00005	2.104	0.0332	0.0354
= Fx 4	0.00023	0.00005	4.692	0.0001	0.0000
= Fx 5	0.00025	0.00003	8.078	0.0001	0.0000
log(Bigr next) * Cluster=Fx1	-0.00037	0.00019	-2.004	0.0394	0.0451
= Fx 2	0.00007	0.00011	0.622	0.5454	0.5337
= Fx 3	0.00016	0.00005	3.289	0.0016	0.001
= Fx 4	-0.00004	0.00004	-0.891	0.382	0.3731
= Fx 5	0.00015	0.00003	5.578	0.0001	0.0000
PhrPos = nonfin * Clust=Fx1	-0.0003	0.00101	-0.276	0.7822	0.7827
= Fx 2	-0.00006	0.0005	-0.128	0.9218	0.8979
= Fx 3	0.00108	0.0002	4.949	0.0001	0.0000
= Fx 4	0.00124	0.00026	4.809	0.0001	0.0000
= Fx 5	0.00006	0.00012	0.528	0.576	0.5973
…					

N	206858	Groups: word	2561
Log likelihood	418926	R^2	0.429386
Deviance	-838425	AIC_c	-837774.5

9 Word category interactions

It was shown in the preceding section that a four-cluster solution (of H-Clust A) produced an optimal solution out of the several hypotheses of word categorizations that were considered. If these word categories are meaningfully distinct, then we have two expectations: [1] that each word category will exhibit differences in reduction patterns overall, and [2] that factors posited to be sensitive to the content and function word division by previous research—including frequency, predictability (e.g., Bell et al. 2009), and phrase position (e.g., Inkelas & Zec 1993; Selkirk 1996)—should also be sensitive to the word categories proposed here. This section presents the model results for the models using the four-cluster solution and examines the interactions involving word clusters. The evidence here and the results presented in Chapter 6 point to the necessity of a four-part division of the lexicon—contra the standard binary division—given the differential behavior of reduction between each of the four word categories. Sections 9.1 and 9.2 present the results of the models for phone distance and duration distance, respectively. Discussion and conclusions of these findings follow in §9.3.

9.1 Interactions in the phone distance model

Table 17 provides the results for the phone distance model. The control, simple fixed effects remain the same as with the control model: all simple effects, as well as the interactions between frequency and predictability, behave as expected, with the sole exception of the onset condition, which was already discussed in §7.2.

In terms of main effects for word cluster, the model results show that Cluster 2, which contains a mix of lexical items (e.g., past tense verbs) and grammatical items (e.g., modal auxiliaries), does not differ significantly from Cluster 1, which contains mostly lexical items (e.g., nouns, adjectives, present tense verbs). Cluster 3 differs significantly from the first two clusters, with more segmental reduction ($\beta = 0.002863$)

exhibited by words in the Cluster 3 category (e.g., prepositions, pronouns, some *wh* words). Cluster 4 also shows no significant main effect differentiating it from the preceding three clusters. However, both Clusters 2 and 4 demonstrate crucial interactions, and thus cannot be discounted on the behavior of the simple effects alone.

The main effect for word frequency indicates that increasing frequency generally signals more reduction. For Cluster 2 words, as compared to Cluster 1 words, the effect is lessened: the model shows that, with increasing frequency, less reduction occurs for Cluster 2 words than for Cluster 1 words (though reduction is still the predicted effect, as shown by the combination of the main effect and interaction slopes: 0.001828 + -0.0002953 = 0.0015327). The other significant interaction with frequency is seen with Cluster 4 words: when compared to the preceding three clusters, Cluster 4 words exhibit an enhanced effect of frequency. Words in Cluster 4 (e.g., determiners, conjunctions) demonstrate additional segmental reduction when they are more frequent. Cluster 3 words appear to exhibit no additional reduction than the previous two clusters with increasing frequency.

Table 17. Four-cluster regression model for phone distance

Factor	Estimate	Std. Error	t value
Intercept	-0.006299	0.0009807	-6.423
log(Speech rate)	0.002641	0.0003913	6.75
log(Word frequency)	0.001829	0.0004239	4.314
log(Bigram previous)	0.001359	0.00005641	24.097
log(Bigram next)	0.001306	0.0000509	25.652
Accent ratio	-0.1581	0.02673	-5.914
Phrase position = nonfinal	0.005447	0.0002238	24.34
Coda = Y	0.01558	0.001128	13.817
Complex coda = Y	0.002641	0.0003778	6.991
Onset = Y	-0.006749	0.001454	-4.641
Word cluster = 2	0.0001095	0.0002609	0.42
= 3	0.002863	0.0002835	10.099
= 4	-0.0001182	0.0003135	-0.377
Interactions			
log(Freq) * log(Bigram prev)	0.0001659	0.00003274	5.066
log(Freq) * log(Bigram next)	0.0003721	0.00002873	12.954
log(Freq) * Cluster = 2	-0.0002953	0.0001344	-2.197
= 3	0.00000021	0.0002065	0.001

= 4	-0.001083	0.0002314	-4.679
log(Bigram prev) * Cluster = 2	0.0002137	0.0000727	2.94
= 3	0.0001301	0.0000439	2.964
= 4	0.0004068	0.00004156	9.789
log(Bigram next) * Cluster = 2	-0.0001997	0.00007171	-2.785
= 3	0.00002207	0.00004296	0.514
= 4	0.0001342	0.00003398	3.951
Phr Pos = nonfin * Cluster = 2	0.002057	0.0002973	6.919
= 3	0.001348	0.0001781	7.572
= 4	0.0001434	0.0001337	1.072
Groups	Name	Variance	Std. Deviation
Word	(Intercept)	0.00014882	0.012199
Residual		0.00100561	0.031711
N	206858	Groups: word	2561
Log likelihood	418991	R^2	0.4292792
Deviance	-838404	AIC_c	-837924.3

For preceding bigram predictability, we see that reduction increases across all three clusters as the probability of the target word given the preceding word increases. This effect, while significant for Clusters 2 and 3, is particularly magnified for Cluster 4, which has the largest coefficient slope (β = 0.0004068) amongst the clusters (cf. β = 0.0002137, 0.0001301). The results of this interaction show that when the probability of the target word given the preceding word increases for words in Cluster 4, the target word demonstrates significantly more reduction than words belonging to the three other clusters.

Results for the interaction with following bigram predictability also demonstrate an increase in reduction correlating with clusters along the stress potential scale. As compared to the other three clusters, Cluster 4 words are predicted to have more reduction when the probability of the target word increases given the following word. The same effect is not significant for Cluster 3 words, when compared to Clusters 1 and 2. Cluster 2 words, when compared to Cluster 1 words, actually demonstrate a slight lessening of the reduction effect of following bigram predictability: predictable Cluster 2 words reduce less than predictable Cluster 1 words.

Phrasal position interacts significantly with the word clusters, with less reduction when the target word occurs in phrase-final position. The results show that Clusters 2 and 3 differ significantly from the preceding clusters in exhibiting more reduction in non-phrase-final positions. The difference between Cluster 4 and the mean of the preceding clusters (1–3) is not significant.

The results of a four-cluster model show that there are crucial differences between how reduction patterns work with respect to the four word clusters. Factors that have been previously shown to be sensitive to content and function word categories are shown here to also be sensitive to the four word clusters. In comparison, the two-cluster model using a standard content versus function word division (see Table 12) is prone to loss of resolution in the interactions. For example, the binary cluster model in Table 12 suggests that an interaction between following bigram predictability and word categories is not significant for a model of segmental reduction. But, we see from the four-cluster model presented in this section that the crucial effect of the interaction between following conditional bigram probability and word categories is limited to words belonging to Cluster 4, which includes the most prototypical of function words. When compared to the other three clusters, Cluster 4 words reduce significantly more when they are more predictable given the following word. Similarly, the two-cluster model indicates that there is less segmental reduction for function words that are more frequent than for content words. While this conclusion from the two-cluster model appears to align with Bell et al.'s (2009) findings that content words are affected by frequency effects but function words are not, the four-cluster model in fact shows that this conclusion is an over-simplification of the reduction facts. In actuality, Cluster 4 words are also sensitive to frequency effects—only Cluster 3 function words demonstrate no additional effect of frequency when compared to Clusters 1 and 2.

While the model results in Table 17 demonstrate that there are differences between the clusters in terms of interactions, how the clusters actually behave with respect to frequency, predictability, and phrase position is more difficult to interpret given the reverse Helmert coding of the clusters in the model (see §8.1). That is, because the model compares the target factor level of the clusters to the mean of all of the previous clusters (e.g., Cluster 4 vs. Clusters 1, 2, 3), the interaction term only reveals whether there is a significant difference and the direction of the difference between the target level and the previous levels. Instead, to examine the actual effect of the factors per each cluster, we look at subset regression for each cluster. The results of each subset regression are given in Table 18 for the relevant factors; full model results are available in Appendix F.

Table 18. Results of subset regressions by cluster (for phone distance)[58, 59]

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Frequency	β	0.0006419	0.002823	0.005323	0.0216536
	*p*mcmc	0.0001	0.0001	0.034	0.003
		***	***	.	*
Bigram pred previous	β	0.0001067	0.0008756	0.001501	0.0042141
	*p*mcmc	0.0008	0.0001	0.0001	0.0001
		**	***	***	***
Bigram pred next	β	0.0002071	0.0007345	0.001852	0.003267
	*p*mcmc	0.0001	0.0001	0.0001	0.0001
		***	***	***	***
Phrase pos = nonfinal	β	0.0009252	0.004946	0.007487	0.0063784
	*p*mcmc	0.0001	0.0001	0.0001	0.0001
		***	***	***	***

. significant at $p < 0.05$, * $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$

The results of the subset regression show that non-final phrase position and increasing unigram frequency and bigram predictability all correlate with more segmental reduction across the clusters, as evidenced by the positive-slope β coefficients. We have already seen from the full model (Table 17) that the magnitude of effect of these factors varies significantly from cluster to cluster.[60] What the subset regression results reveal

---

[58] Because *p*MCMC values are considered to be more conservative (Baayen 2013:97; a.o.), I do not report *p*-values based on *t*-tests in this table. The results are fairly similar, except for some small differences where *p*MCMC demonstrates marginal significance and *Pr* (>|*t*|) demonstrates marginal insignificance, or vice versa. See Appendix F for the full model results, including *t*-test-based *p*-values.

[59] Interactions between frequency and bigram predictability (both previous and following) are not shown here but are significant for all of the subset models, with the exceptions of (1) freq * prev bigram prob for the Cluster 3 subset model and (2) marginal significance for freq * prev bigram prob for the Cluster 1 subset model.

[60] It is also important to note that direct comparisons between β coefficients cannot be made with subset regression models due to differing n.

is that the effects—with the exception of frequency—are significant within each cluster, even if the full model does not show them to be significantly different when compared to the group of preceding clusters. For example, Helmert coding obscures the root cause of the insignificance of the interaction between Cluster 4 and phrase position. It is unclear whether this insignificance stems from phrase position having no role in predicting segmental reduction for Cluster 4 words, which runs counter to our theoretical expectations (see §2; Inkelas & Zec 1993; Selkirk 1996; a.o.), or if this insignificance arises from Cluster 4 being insufficiently distinct from Clusters 1–3 in terms of phrase position effects, which still remain significant within-cluster. The results of the subset regression model clarifies this issue, demonstrating that phrase position significantly predicts segmental reduction within Cluster 4 words. Thus, we can conclude from the full model that the effect of phrase position for Cluster 4 words is not distinct from the effect of phrase positions for the preceding Clusters 1–3 together.

In other cases, the results of the subset regression models parallel the interaction results from the full model. The subset regressions show that frequency is less reliable of a predictor for Clusters 3 and 4 than it is for Clusters 1 and 2; in particular, the effect of frequency is only marginally significant for Cluster 3 words. The full model also showed an insignificant effect of a Cluster 3 and frequency interaction. The subset regression suggests that this insignificance is based in part on the fact that frequency does not have a strong effect in predicting segmental reduction in Cluster 3 words. We cannot completely discount frequency effects in Clusters 3 and 4, however, as most of the interactions between frequency and previous/following bigram predictability remain highly significant for Clusters 3 and 4 in subset regressions (see Appendix F). The only interaction that does not achieve significance in the models is the interaction between frequency and previous bigram predictability for Cluster 3 words.

Due to Helmert coding and the nature of multi-level categorical predictors in regression models, the full model also makes it difficult to see differences (and similarities) between individual word categories. For example, the full model indicates that Cluster 4 words behave significantly differently from the average of Clusters 1–3 words with respect to the influence of word frequency on reduction. From the full model, however, it remains unclear whether this significant difference is between Clusters 4 and a single preceding level (e.g., Cluster 3) or between Cluster 4 and the composite behavior of the preceding word categories. Essentially, it is impossible to tell from the full model how different clusters are from each other. Understanding individual pairwise relationships between clusters serves to elucidate how distinct each proposed word category is from the others. Pairwise relationships are examined via pairwise subset regressions, which compare the reduction behavior of each word category to every other category (e.g., Cluster 1 versus Cluster 2; Cluster 2 versus Cluster 3; etc.). A summary of the results using *p*MCMC values for interaction terms in pairwise comparison models of phone distance is given in Table 19.

The results of the pairwise comparisons demonstrate that some cluster pairs are more distinct from each other than from others. For example, Cluster 4 demonstrates significant differences in reduction sensitivity to frequency, bigram predictability, and phrase position from Clusters 1 and 2. The difference between Cluster 3 and 4 is more subtle, differing significantly only in behavior with respect to preceding bigram probabilities and phrase position. The pairwise results also show that not all clusters differ from others in the same ways. We see significant differences in the interactions of phrase position and clusters almost across-the-board, whereas there is less distinction between the clusters in frequency and predictability effects. We will return to a discussion of these results in §9.3.

*Table 19. p*MCMC-values for interactions in pairwise subset regressions (for phone distance)

		Cluster 2	Cluster 3	Cluster 4
Cluster 1	Frequency	0.0001 ***	0.0738	0.0038 *
	Bigr pred prev	0.0001 ***	0.0001 ***	0.0001 ***
	Bigr pred next	0.5272	0.6978	0.0001 ***
	Phrase pos=nf	0.0001 ***	0.0001 ***	0.0001 ***
Cluster 2	Frequency		0.0832	0.0014 *
	Bigr pred prev		0.165	0.0001 ***
	Bigr pred next		0.6644	0.0001 ***
	Phrase pos=nf		0.0026 *	0.5146
Cluster 3	Frequency			0.1122

Bigr pred prev	0.0001 ***
Bigr pred next	0.3604
Phrase pos=nf	0.0036 *
. significant at $p < 0.05$, * $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$	

9.2 Interactions in the duration distance model

Results for duration distance are given in Table 20. Like the phone distance model results in Table 17, the simple effects of word cluster show a significant difference between Clusters 3 and preceding Clusters 1–2, with more reduction exhibited by clusters that contain a greater number of prototypical function words and fewer prototypical lexical words. While Clusters 2 and 4 show no significant simple effects, there are interactions to consider. The interactions for the duration distance model are largely similar to the results of the phone distance model. The two primary differences include [1] the interaction between frequency and the word categories and [2] the behavior of Cluster 4 with respect to phrase position. The interaction effects of bigram predictability (both previous and following) are the same as with phone distance, so they will not be presented again here.

As with the phone distance model, the interaction between unigram frequency and word clusters is not significant for either Clusters 2 or 3. For Cluster 4, there appears to be a curious effect of less reduction that is correlated with increasing frequency, which demonstrates marginal significance. Subset regression models (Table 21) provide better understanding of the interaction of frequency and word cluster. The results of subset regressions show that frequency is an unreliable factor in the prediction of duration reduction for words in Clusters 3 and 4. This pattern is similar to what was found with segmental reduction, in that Clusters 3 and 4 also demonstrated diminished effects of frequency, as compared to Clusters 1 and 2. Like the phone distance subset regressions, frequency does interact significantly with bigram predictability even for Clusters 3 and 4 (i.e., higher frequency, more contextually-predictable words lead to even more reduction) and cannot be completely dropped as unimportant from the models. The subset regression results also shows that frequency behaves as expected for words in Clusters 1 and 2: more frequent words in these clusters correlate with shorter durations from their estimated citation durations. The results of the full model as far as the interaction between frequency and Cluster 2 can thus be interpreted as Cluster 2 words being insufficiently distinct in frequency behavior from Cluster 1 words.

Table 20. Four-cluster regression model for duration distance

Factor	Estimate	Std. Error	t value
Intercept	0.0001506	0.00009543	1.58
log(Speech rate)	0.002012	0.00003886	51.78
log(Word frequency)	0.0002525	0.00004158	6.07
log(Bigram previous)	0.0001938	0.0000056	34.59
log(Bigram next)	0.0002615	0.00000505	51.74
Accent ratio	-0.006257	0.002604	-2.40
Phrase position = nonfinal	0.003017	0.00002222	135.77
Coda = Y	0.0007346	0.0001106	6.64
Complex coda = Y	0.0001401	0.00003741	3.75
Onset = Y	0.0007946	0.0001421	5.59
Word cluster = 2	-0.000003289	0.00002588	-0.13
= 3	0.0002113	0.00002806	7.53
= 4	-0.00004506	0.00003088	-1.46
Interactions			
log(Freq) * log(Bigram prev)	0.00002476	0.000003248	7.62
log(Freq) * log(Bigram next)	0.00006137	0.000002849	21.54
log(Freq) * Cluster = 2	-0.00006137	0.0000133	-1.98
= 3	-0.0000664	0.0000203	-0.33
= 4	-0.00005331	0.00002274	-2.34
log(Bigram prev) * Cluster = 2	-0.00000007	0.0000072	-0.01
= 3	0.00001352	0.00000436	3.10
= 4	0.00001367	0.000004127	3.31
log(Bigram next) * Cluster = 2	-0.00004221	0.000007119	-5.93
= 3	0.000003292	0.000004266	0.77
= 4	0.00001714	0.000003374	5.08
Phr Pos = nonfin * Cluster = 2	0.0001011	0.00002952	3.42
= 3	0.00008718	0.00002952	4.93
= 4	0.00005934	0.00001328	4.47
Groups	Name	Variance	Std. Deviation
Word	(Intercept)	0.0000014	0.0011842
Residual		0.0000099	0.0031495
N	206858	Groups: word	2561
Log likelihood	896680	R^2	0.2785268

Deviance	-1793906	AIC$_c$	-1793301

Table 21. Results of subset regressions by cluster (for duration distance)[61]

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Frequency	β *p*mcmc	0.0001703 0.0001 ***	0.000197 0.0001 ***	0.00006454 0.7447	-0.0001204 0.821
Bigram pred previous	β *p*mcmc	0.00008609 0.0001 ***	0.0001312 0.0001 ***	0.0003128 0.0001 ***	0.0003722 0.0001 ***
Bigram pred next	β *p*mcmc	0.0001106 0.0001 ***	0.0001643 0.0001 ***	0.0004317 0.0001 ***	0.0005421 0.0001 ***
Phrase pos = nonfinal	β *p*mcmc	0.002324 0.0001 ***	0.002348 0.0001 ***	0.003576 0.0001 ***	0.003206 0.0001 ***

. significant at $p < 0.05$, * $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$

The other somewhat minor difference between the phone distance and duration distance results is that we see, for duration distance, a consistent increase in the effect of phrase position on durational reduction. Words in non-final positions are more likely to exhibit shorter durations (and, likewise, words in phrase-final position correlate with longer durations). The full model demonstrates that this effect increases significantly as we move along the scale of reducibility from Cluster 1 to Cluster 4 words. In the phone distance model by comparison, the difference between Cluster 4 words and words of the preceding clusters does not reach significance ($t = 1.072$) but the effect remains the same.

Pairwise subset regression results are provided in Table 22. The results are more or less similar to results predicting segmental reduction; differences and similarities are discussed in the following section.

[61] Interactions between frequency and bigram predictability (both previous and following) are not shown here but are significant for all of the subset models without exception.

Table 22. pMCMC-values for interactions in pairwise subset regressions (for duration distance)

		Cluster 2	Cluster 3	Cluster 4
Cluster 1	Frequency	0.0366 .	0.0002 **	0.0001 ***
	Bigr pred prev	0.8066	0.0006 **	0.0002 **
	Bigr pred next	0.0006 **	0.138	0.0001 ***
	Phrase pos=nf	0.0001 ***	0.0001 ***	0.0001 ***
Cluster 2	Frequency		0.4996	0.0004 **
	Bigr pred prev		0.0078 *	0.0068 *
	Bigr pred next		0.2168	0.0001 ***
	Phrase pos=nf		0.0168 .	0.0001 ***
Cluster 3	Frequency			0.0076 *
	Bigr pred prev			0.7486
	Bigr pred next			0.4384
	Phrase pos=nf			0.3022
. significant at $p < 0.05$, * $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$				

9.3 Discussion of interactions

To summarize, the results in this section show that factors previously identified to be sensitive to content and function word categories demonstrate further sensitivity to the four-cluster word categorization posited in previous sections. Each word class displays significant effects of reduction correlating with frequency, conditional bigram probability, and phrase position in the expected directions. More frequent and predictable items correlate with more phonetic reduction, and phrase-final words are prone to less reduction than non-phrase-final words. These results are summarized in Table 23.

Table 23. Summary of results: Individual effects

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Frequency	✓✓✓	✓✓✓	✓ (*P*)	✓✓ (*P*)
Bigram previous	✓✓✓	✓✓✓	✓✓✓	✓✓✓
Bigram next	✓✓✓	✓✓✓	✓✓✓	✓✓✓
Phrase position	✓✓✓	✓✓✓	✓✓✓	✓✓✓
✗ not significant, ✓ marginally significant, ✓✓ significant, ✓✓✓ highly significant				
P = significant only for phone distance, *D* = significant only for duration distance.				

Moreover, it was shown that frequency, predictability, and phrase position effects on word pronunciation are dependent on the presence of differential word categories, with significant differences for each of the four word clusters. These results, which are summarized in Table 24, mean that patterns in reduction rely on clear categorical distinctions of lexical category: the effects for frequency, predictability, and structural factors modulate based on the category of the word. Generally, clusters with more prototypical function words (Clusters 3 and 4) show enhanced effects of reduction in predictable and phrase-internal environments.

Table 24. Summary of results: Differences between clusters

		Cluster 2	Cluster 3	Cluster 4
Cluster 1	Frequency	✓✓✓	✓✓✓ (*D*)	✓✓✓
	Bigr pred prev	✓✓✓ (*P*)	✓✓✓	✓✓✓
	Bigr pred next	✓✓✓ (*D*)	×	✓✓✓
	Phrase pos=nf	✓✓✓	✓✓✓	✓✓✓
Cluster 2	Frequency		×	✓✓✓
	Bigr pred prev		✓✓ (*D*)	✓✓✓
	Bigr pred next		×	✓✓✓
	Phrase pos=nf		✓✓	✓✓✓ (*D*)
Cluster 3	Frequency			✓✓ (*D*)
	Bigr pred prev			✓✓✓ (*P*)
	Bigr pred next			×
	Phrase pos=nf			✓✓ (*D*)

× not significant, ✓ marginally significant, ✓✓ significant, ✓✓✓ highly significant

P = significant only for phone distance, *D* = significant only for duration distance.

We can compare the results generated here to conclusions from previous studies of reduction in content and function words. In particular, previous work by Bell et al. (2009) on duration reduction in conversational data finds that frequency and predictability are sensitive to word categories in that the effects of these predictors are only significant for some word categories and not all. A table summarizing their findings is reproduced below.

Table 25. Summary of Bell et al.'s results (2009:104)

	Content words	Mid/low-frequency function words	High-frequency function words
Frequency	✓✓✓	✗	✗
Bigram previous	✗	✓	✓✓✓
Bigram next	✓✓✓	✓✓✓	✗

✗ not significant, ✓ marginally significant, ✓✓ significant, ✓✓✓ highly significant

For frequency, Bell et al. find that only content words appear to be sensitive to frequency effects, with no significant increase of reduction correlated with more frequent function words. The duration results of the current study are similar. In terms of durational reduction, Clusters 1 and 2, which contain standard content word classes, have significant effects of frequency: more frequent words exhibit more reduction than less frequent words. The effect is especially strong for Cluster 1 words and is then significantly lessened for words belonging to Cluster 2. In contrast, Clusters 3 and 4 show no significant effects for frequency in individual subset regressions.

Segmental reduction, however, tells a slightly different story for frequency. Words of Clusters 3 and 4 demonstrate significant correlations between frequency and the amount of reduction, indicating that more frequent words in these clusters will exhibit more segmental reduction, just as more frequent Cluster 1 and 2 words do. However, from the results of the full model and the comparative subset regressions, we know that the effect of frequency for Cluster 3 and 4 words is much less than the effect for words in Clusters 1 and 2 (especially Cluster 1). The segmental reduction results for frequency suggest that frequency is a much more important and reliable predictor of reduction in prototypical content words than it is for all other words, but that the effect is still present for all words.

The issue of frequency's interaction with word class is part of a long-standing debate on whether there are differential effects between function and content words in their sensitivity to frequency. Bell et al.'s results appear to align with one of the earliest studies on the issue, Bradley (1978; et seq.), which reported no sensitivity from function words to frequency in a lexical decision task. Based on their results, Bell et al. argue for a psycholinguistic model of language production in which content words are part of a slower mode of access that results in synchronization mismatches in frequency-based activation and articulation planning (2009:106). Under this model, function words form their own level or mode of access in production in which frequency has no affect.

Subsequent research after Bradley (1978), however, has reported opposite results that counter Bradley's and Bell et al.'s. Based on a range of evidence from lexical decision tasks to neurological studies, this line of research shows instead that frequency does affect both function and content words (Gordon & Caramazza 1982; Segui et al. 1982; Segalowitz & Lane 2000; a.o.). Segalowitz & Lane (2000) report that function words have a lessened sensitivity to frequency effects, but this difference in word categories can be attributed to contextual predictability and frequency differences between function and content words (e.g., high frequency function words show no effect of frequency for reduction, but mid- and low frequency function words do). Because frequency effects are found for both function and content words, Segalowitz & Lane argue for a model of language production in which access for function words are "privileged" but are not separate from content words, as is proposed by Bell et al.

The segmental reduction results presented in the current study demonstrate that frequency does still play a role in predicting reduction for function words—albeit this is a lesser role than the one it plays for content words. Unlike Bell et al.'s findings, it is not the case here that there is a complete lack of frequency effect for function words: it is just that the effect is smaller and only observable in the segmental reduction condition where the differences between word categories are more pronounced. Such a result raises questions for a strong interpretation of Bell et al.'s dual access-based explanation for why content and function words differ in their sensitivity to frequency. Namely, if function words are supposedly accessed via a mode that does not involve frequency effects, how can the frequency effects that we find here—and that Segalowitz & Lane (2000) show—be explained? It appears instead that the facts here point to an access and production model that allows frequency to affect both content and function word pronunciations, and that the differential behavior of frequency in “privileged access” with respect to word categories must arise from some other explanatory source.

One such source, which the previous literature on the topic leaves largely unexplored, is that the difference between word categories in frequency interaction stems from the unbalanced category membership between word clusters. While the number of word tokens in each cluster is similar at about 45,000 words per cluster, the number of word types in each cluster is vastly different. Cluster 1 consists of 2,133 unique word types. Cluster 2 has a lesser membership of 640 unique word times. Clusters 3 and 4 have drastically fewer word types: 94 and 24, respectively. Thus, once the effect of individual words is accounted for via the random effect of word, the leftover effect of frequency is expected to be smaller for the clusters that have a sparser frequency distribution to explain the same amount of data. Under such an explanation, we can see why the frequency effect may decay from one word cluster to another, as unique word types grow smaller per category. As such, “privileged access” of more function-like words stems from within-category sparseness. Similar claims of within-category sparseness have been made to explain privileged access of affix versus root morphology (Ussishkin & Wedel 2002; 2009).

10 Interim conclusion

This chapter has revisited the empirical phonetic evidence underlying the basic assumption behind word categories: that function words are more likely to reduce than content words. Results based on a study of reduction in conversational American English demonstrate that a simple, oft-assumed content versus function word division does not as accurately model the phonetic patterns as a more fine-grained, four-way division. This result parallels previous findings based on pitch accent placement in spoken American English (e.g., Hirschberg 1993) and provides principled, quantifiable evidence of stress differences between lexical categories.

The optimal four-way categorization presented in §8.3ff largely mirrors standard divisions of content and function word categories, but also allows for more subtlety in dividing up the lexicon. Impressionistically, the differences between previous assumptions of the content/function division and the results presented here align with broader expectations. For example, the grammatical words that reduce less often and pattern more like lexical words are the ones that appear to carry more semantic information or are more closely tied to the verb complex (e.g., auxiliaries).[62] This propensity for reduction also reflects common paths of grammaticalization from content-ful to functional elements (e.g., Hopper & Traugott 1993; Liu et al. 2010), and relates to reduction patterns expected by a Smooth Signal or Uniform Information Density view on speech stream fluctuations (Aylett & Turk 2004; Frank & Jaeger 2008; Jaeger 2010; a.o.). Future work examining the influence of vowel quality (i.e., whether the vowel gets centralized as a proxy for reduction), for example, could continue to illuminate the relationship between reduction and predictability (e.g., Cutler 1993).

A four-way division of the lexicon raises some question about how such a division is built into formal models, which generally assume a binary division between *Fnc* and *Lex* elements (cf., Anttila 2012). Having four categories, for instance, requires reconsideration of how certain words—traditionally considered the function word set—could remain invisible from prosodic and phonological computation (e.g., Principle of Categorial Invisibility from Selkirk 1984) while other function words appear to have a certain level of underlying stress that reduces their likelihood of surface phonetic reduction. Similarly, it is up for grabs how the systematic phonetic differences between certain content words should be modeled as well, if the theoretical assumption has been that content words across-the-board are protected from reduction. As a caveat, it should be noted here that the model presented in this chapter has yet to narrowly control for prosodic and syntactic environments, so it is still a possibility that structural distribution can further simplify the underlying division of the lexicon back to a binary one. However, the results here remain robust even after bigram predictability, frequency, rough phrasal position, and individual word have been accounted for.

[62] It is left to future research to examine this issue further.

One open question for future consideration is the place of secondary stress in polysyllabic words along the scale of reducibility. Inkelas & Zec (1995:211) hypothesize that the stress of function words "is interpreted at best as secondary prominence." Thus, under this hypothesis, we would expect that secondary stress in polysyllabic words would mirror the reduction patterns of certain function elements rather than the reduction pattern of primary stress of content elements. An alternative outcome would be that secondary stress patterns independently. Conclusions from this future work will inform the issue of how these differences in lexical stress are encoded in the lexicon or our models of grammar.

The next chapter turns to rhythmic considerations arising specifically from the lexical categories uncovered here.

Chapter 6. Rhythm and reduction

1 Introduction

The supposition that there are crucial difference in lexical stress for content versus function words allows us to consider how such differences interact with the Principle of Rhythmic Alternation. Recall that there are a number of compensation strategies for satisfying rhythmic regularity (see Chapters 1 and 2), falling into three broad categories:

1. Phonetic adjustment (e.g., reduction/enhancement of the phonetic signal)
2. Phonological adjustment (e.g., stress shift, foot flipping)
3. Morphosyntactic adjustment (e.g., suppletion, ordering, periphrasis)

Following the assumption that function words are eligible for reduction but content words are not (Selkirk 1984; 1996; Inkelas & Zec 1993; a.o.) (see also Chapter 5), phonetic adjustments for satisfying rhythmic regularity should only be available where function words are concerned. Content words, in contrast, should be blocked from such phonetic poorly-formed rhythmic patterns because they do not undergo reduction; therefore, sequences involving content words need to look elsewhere—that is, to phonological or extra-phonological adjustments—to achieve rhythmicity. This is to say that once we assume that stress properties are not equal across classes of words, then the mechanisms by which rhythmic regularity is achieved should also reflect such lexical, stress-based inequalities.

The conclusions of the previous chapter complicate this issue further. The study of word category reduction in Chapter 5 concluded that, based on reduction patterns in spoken language data, there are four major classes of words that exhibit differential behaviors with respect to how often and how much they reduce. If it is the case that there are more than two word categories along a content-to-function word spectrum, then such categories should also demonstrate different reactions in the face of non-optimal rhythm. More prototypical function words (e.g., Cluster 4 words), which have more reducibility potential, should have greater access to phonetic adjustments for rhythmic irregularity than prototypical content words (e.g., Cluster 1 words), which we know to have less phonetic reducibility potential. This chapter examines the interaction between rhythmic environment and phonetic reduction for the word categories proposed in Chapter 5, focusing in particular on the ability for words to reduce when in stress clash environments. The results crucially show that sensitivity to stress clash differs amongst the word categories: words that have more reducibility potential are the ones that display greater phonetic adjustments when neighboring syllables are stressed.

The results presented here return us to key questions about the availability of phonetic, phonological, and morphosyntactic information and the behavior of rhythm in syntagmatic contexts across word boundaries. One issue that arises concerns the availability of phonetic, phonological, and morphosyntactic adjustments for rhythm at various stages of grammar and language production. If it is the case that different words are prone to different compensation strategies, then how do these compensation strategies interact? Are they all available at once for the speaker in computing rhythmic regularity, or does there exist some opacity between syntactic and phonetic information, as is commonly assumed?

Another key issue that remains open is the question of how these word categories are built into the mental lexicon. Previous theories of word categorization have put forth at least two conceptualizations of word class differences: [1] that word categories (e.g., content, function) comprise separate lexicons (or utilize diacritic representations of *Lex* and *Fnc*), or [2] that word categories directly encode underlying differences as lexical stress (i.e., content words have underlying lexical stress and function words do not). The word categorization results presented here has been thus far non-committal regarding lexical encoding. On one hand, reducibility potential could be construed as a diacritic for separate sections of the mental lexicon—or separate lexicons completely; or, on the other hand, reducibility potential could be directly built in as levels of stress in underlying word representations. While the predictions of these two theories share a fair amount of overlap, there are some differences in predictions, in that direct stress in underlying lexical representations should produce direct implications for operations that are sensitive to underlying stress information. Conversely, a theory based on lexical indexation does not make such direct predictions.

This chapter offers preliminary explorations of these two outstanding issues as they relate to syntactic choice. (A full study of these issues goes beyond the scope of the current thesis and is not attempted here.) Returning to the English genitive alternation (see Chapter 4), the latter half of this chapter addresses the ramifications of the word category results on rhythm in a syntactic alternation. The findings suggest that ordering between phonetic repairs and extra-phonological repairs (e.g., construction choice) must still be maintained: it is argued that rhythmic computation for the purposes of syntactic choice demonstrates little or no look-ahead knowledge of the possibility of phonetic repairs for poor rhythmicity. Finally, patterns of end weight preferences in genitive construction choice indicate that scalar word category information may go beyond reducibility potential, and to some limited extent, the differences in word categories may be encoded in underlying lexical stress representations, to which the morphosyntactic component has access.

The chapter is organized as follows. Section 2 investigates the differences in phonetic reduction by word category in non-optimal rhythmic environments. Section 3 revisits the English genitive alternation in light of the newly proposed content and function word classification system: this section addresses the issues of lexical stress representation and the interaction between rhythmic compensation strategies via two exploratory studies of rhythm and end weight in the genitive construction (§§3.1 and 3.2, respectively). Section 4 concludes, with discussion of what the results presented herein reveal about the architecture of the linguistic system.

2 Word category differences for rhythm

In this section, we examine the effect of rhythmic environment on reduction in word categories.

2.1 Predictors

The current study builds on the models presented in the previous chapter to investigate the reduction behavior of word categories in rhythmic environments. Newly added predictors are discussed in turn.

2.1.1 Rhythmic environment

This study looks specifically at stress clash avoidance and phonetic reduction. Following the Principle of Rhythmic Alternation, the expectation is that phonetic adjustments can be used to resolve stress clash. Neighboring stressed syllables should therefore cause more reduction, and we expect this effect to be stronger on target syllables that have a greater reducibility.

Because the Principle of Rhythmic Alternation also assumes that rhythmic well-formedness is assessed across strings of words, we furthermore might expect there to be an interaction between preceding and following stress clash environments. If it is the case that a monosyllabic word has a high stress potential but is surrounded on both sides by strong syllables, then reduction—and even more of it—should occur in order to avoid stress clash with the surrounding syllables. If it is the case that a monosyllabic word with a low stress potential is surrounded on both sides by strong syllables, then less reduction should occur to prevent the surrounding syllables from clashing with one another.

Taken here is a narrow view of stress clash environments: only stressed syllables of polysyllabic words (both primary and secondary stress) or monosyllabic Cluster 1 words (nouns, adjectives, some verbs) were considered to be potential syllables that trigger stress clash. All other environments were coded as not causing stress clash under the assumption that Cluster 1 words have the least reducibility potential and thus are the most likely to induce stress clash violations.

2.1.2 Other predictors

The base models from Chapter 5 were included here as controls. All control predictors remain the same as described in the previous chapter. The only difference is the addition of a phrase-initial control, discussed below.

PHRASE-INITIAL. It is necessary here to add a control for phrase-initial domain because, without preceding phonetic material, these environments cannot result in a stress clash. Phrase-initial environments were not included as a control in the previous chapter because the bulk of the literature on phrase position and reduction focuses primarily on phrase-final position. However, there may be some reasons to expect phrase-initial effects on reduction: for example, Fougeron & Keating (1997; et seq.) show that articulatory strengthening occurs at phrase-domain edges.

2.2 Results and discussion

Results from a phone distance model with stress clash environment are given in Table 17.[63] Results from duration reduction are similar.

[63] For ease of readability, some control predictors not of interest have been excluded from this table.

Table 26. Four-cluster regression model for phone distance

Factor	Estimate	Std. Error	t value
Intercept	-0.006184	0.0009813	-6.302
Phrase position = nonfinal	0.005487	0.0002279	24.078
Phrase position = noninitial	-0.002475	0.0002698	-9.174
Previous stress clash = N	-0.0000608	0.0001983	-0.307
Following stress clash = N	-0.0002765	0.0001834	-1.508
Word cluster = 2	-0.00005141	0.0002667	-0.193
= 3	0.002832	0.0002841	9.967
= 4	-0.0001991	0.0003139	-0.634
…			
Interactions			
Prev stress cl = N * Cluster = 2	-0.0000513	0.0002804	-0.183
= 3	-0.0005458	0.0001445	-3.077
= 4	0.0007575	0.0001254	6.04
Foll stress cl = N * Cluster = 2	-0.0003476	0.000276	-1.259
= 3	0.00005089	0.0001413	0.36
= 4	-0.0002742	0.000104	-2.636
Phr Pos = noninit * Cluster = 2	0.001962	0.000303	6.475
= 3	0.001269	0.0001811	7.009
= 4	0.0001643	0.0001369	1.2
Phr Pos = nonfin * Cluster = 2	-0.00111	0.0001751	-9.253
= 3	-0.00162	0.0001751	-9.253
= 4	-0.001192	0.0001677	-7.108
…			
Groups	Name	Variance	Std. Deviation
Word	(Intercept)	0.00014886	0.012201
Residual		0.00100390	0.031684
N	206858	Groups: word	2561
Log likelihood	419080	R^2	0.4302834
Deviance	-838766	AIC_c	-838077.8

In general, the inclusion of stress clash environment as a predictor of reduction significantly improves the explanatory power of the models, suggesting that there is phonetic accommodation given neighboring rhythmic environments in connected speech. However, it is crucial to note that stress clash does not contribute as much to the explanatory power of the model as many of the other predictors. Drop-one log-likelihood and AIC_c tests of both the full models and individual subset regressions reveal that, though significant, both preceding and following stress clash predictors tend to rank low in terms of contribution to the model. Structural factors, speech rate, and predictability (including unigram frequency and bigram probabilities) consistently make the largest contributions to predicting reduction.

Of greater interest here are the potential differences between word categories in their sensitivity to rhythm and stress clash and whether these differing sensitivities are reflected in the use of phonetic adjustment to resolve non-optimal rhythm. The full model presented in Table 17 shows that there are significant differences between certain word categories in their interactions with preceding and following stress clash environments. Clusters 1 and 2 appear to exhibit no significant differences in their reduction behavior (i.e., previous stress t = -0.183; following stress t = -1.259). The greater differences arise when we look at the clusters that primarily contain prototypical function words: Cluster 3 versus Clusters 1–2 and Cluster 4 versus Clusters 1–3. As is expected, these clusters exhibit significantly more sensitivity to stress clash environments than clusters that contain prototypical content words and function words that do not have high reducibility potentials. Because of the higher reducibility potential for words in Clusters 3 and 4, we expect that they will have less phonetic faithfulness to target, full pronunciations when stress clash is encountered. Conversely, words with low reducibility potential in Clusters 1 and 2 should not vary as widely as to their prosodification.

We can examine the segmental and durational behavior of the word categories with respect to rhythmic environment via individual subset regressions. Results for individual subset models for phone and duration distances are given in Table 27 and Table 28, respectively.

Table 27. Results of subset regressions by cluster (for phone distance)

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Previous stress clash = N	β	-0.000115	-0.0002391	-0.001859	0.0023124
	*p*mcmc	0.4262	0.4388	0.0001 ***	0.0001 ***
Following stress clash = N	β	0.0002687	-0.0004586	0.00005783	-0.0013081
	*p*mcmc	0.0758	0.0912	0.8688	0.0036 *
. significant at $p < 0.05$, * $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$					

Table 28. Results of subset regressions by cluster (for duration distance)

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Previous stress clash = N	β	-0.0000883	-0.0000765	-0.0002123	0.00007009
	*p*mcmc	0.0002 **	0.0022 **	0.0001 ***	0.1182
Following stress clash = N	β	-0.00007248	-0.00006539	0.00000297	-0.0002372
	*p*mcmc	0.0034 **	0.0048 **	0.9242	0.0001 ***
. significant at $p < 0.05$, * $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$					

Across both the duration and phone distance conditions, Clusters 1 and 2 demonstrate little difference to each other with respect to stress-induced reduction. Neither cluster exhibits significant effects of segmental reduction in the face of stress clash, but, in terms of duration, the words in these two clusters tend to be durationally shorter from their estimated target citation forms when there is a neighboring stressed syllable. This significant effect is illustrated in Figure 26.

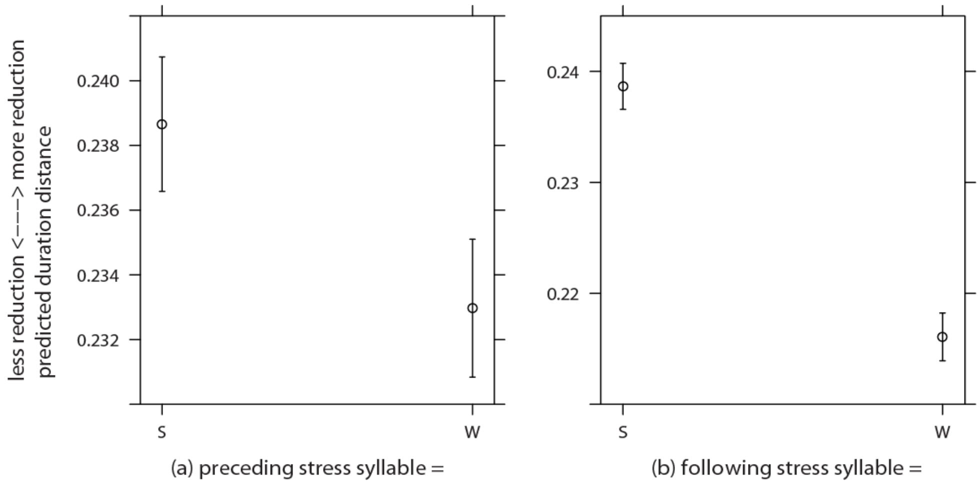

Figure 26. Partial effects plots for stress clash environments in a duration distance model of Cluster 1 words.

Unlike Clusters 1 and 2, the behaviors of Clusters 3 and 4 demonstrate significant differences both from each other as well as from Clusters 1 and 2. Figures 27 and 28 illustrate the differences in both the effects and magnitudes of effects of reduction in stress clash environments

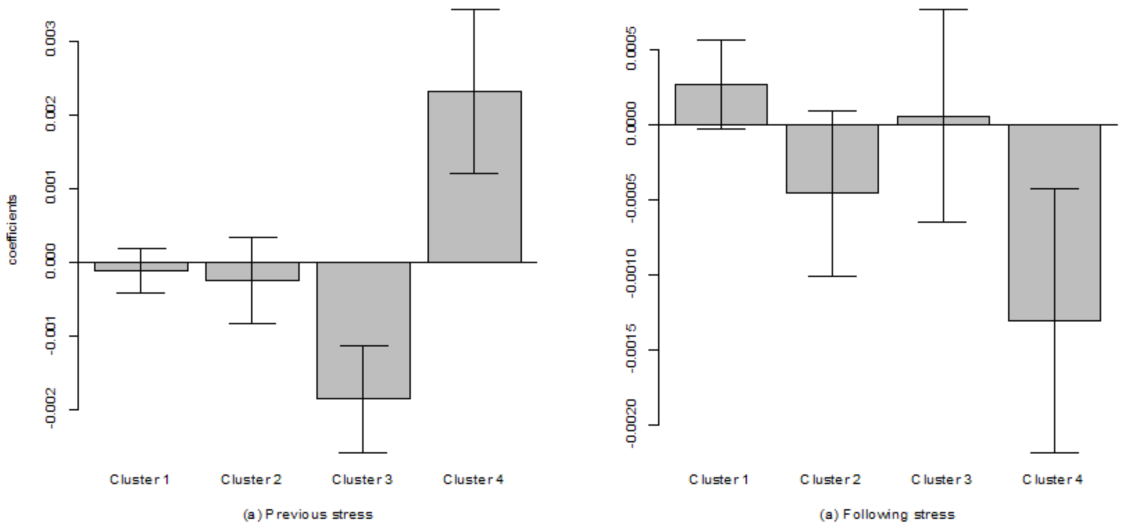

Figure 27. Coefficients and standard errors of stress environment by cluster for phone distance. (Negative indicates more reduction when there is a surrounding stress clash environment; positive indicates less reduction when there is a surrounding stress clash environment.)

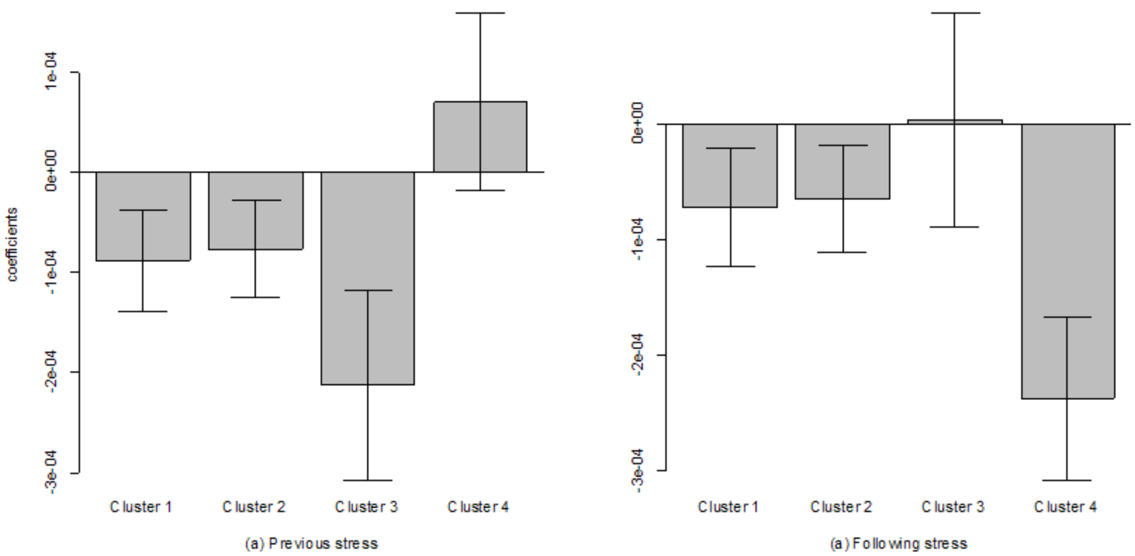

Figure 28. Coefficients and standard errors of stress environment by cluster for duration distance.

Of note is that, when there is a significant effect of stress clash on reduction for words in Clusters 3 and 4, the magnitude of the effect is much larger than any significant effects of rhythmic environment for words in Clusters 1 and 2. This is as we expect, given that words in Clusters 3 and 4 have greater reducibility potentials and are more likely to occur as unfaithful pronunciations of their citation forms.

For Cluster 3 words, we see that there is a reliable effect of preceding stress clash on reduction: words in this category reduce more when the preceding syllable is stressed. There is, however, no effect of following stress clash environment.

For Cluster 4 words, almost all conditions significantly affect reduction, with the exception of the effect of preceding stress clash on durational reduction. When followed by a stressed syllable, Cluster 4 words reduce more, and the magnitude of this effect is significantly greater than the same effect for words of Clusters 1 and 2 in terms of duration distance. Preceding stress clash environment produces different results for words of Cluster 4. The results of the phone distance model (see Figure a) suggest that a preceding stress clash environment correlates with significantly less reduction, which runs counter to the rest of the results—that more reduction should correlate with stressed neighbors. Such an anomaly can be explained by the fact that here we are only investigating immediately adjacent stress clash, and rhythmic patterning involving a wider neighboring environment (e.g., lapse) could also affect reduction, which is not explored here. In particular for Cluster 4 words, there may be a tendency towards parsing in triple rhythm sequences: for example, if the article *the* in *for the love of the game* reduces, *for the* can be parsed together, whereas such additional reduction for a word with less reducibility potential (e.g., *their* in *for their love of their game*) seems much more unlikely (see Selkirk 1984; a.o. for a parallel discussion). An alternative explanation for this result is that Cluster 4 function words—ones with already-low stress potentials—resist reduction when the immediately adjacent syllables are stressed because that would bring the stressed syllables closer to each other and cause stress clash. Reducing Cluster 4 words (the weakest ones already) would amount, in extreme cases, to deletion (or near-deletion), which would result in adjacent stressed syllables. This explanation, however, does not completely provide for the differences observed between previous and following stress environments. Understanding the details of this result requires consideration of more sophisticated rhythmic patterns and the role of footing than is included in the current study; this is saved for future research.

No significant interaction of preceding and following rhythmic environments was found, and thus the interaction was dropped from the final models reported here. For phone distance, stress clash at both boundaries of the target word did not contribute significantly to improvement in the explanatory power of the full model (χ^2 = 8.9463, df = 4, p = 0.06245, $\Delta \mathrm{AIC}_c$ = 53.9); nor did the interaction contribute significantly to the full model of duration distance (χ^2 = 4.3989, df = 4, p = 0.3547, $\Delta \mathrm{AIC}_c$ = 77). Tested in separate individual subset regressions, the only word category in which an interaction between surrounding stress clash environments proved to be even marginally significant and reliable was Cluster 1 words in a model of phone distance (β = 0.000856, S.E. = 0.0003474, t = 2.464, χ^2 = 6.0741, p = 0.01372). However, counter to the expectation that stress clash should result in more reduction, the interaction results indicate that less reduction is actually seen in Cluster 1 words when surrounded by stress clash environments on both edges. Note that there is no significant simple effect of stress clash on either edge of Cluster 1 words. Because the interaction between preceding and following stress clash was not important to the models, the interaction was dropped from the final models reported here.

The finding that there is no interaction in preceding and following word stress on reduction contradicts our expectations about rhythmic patterning, given the Principle of Rhythmic Alternation. There are a few possible explanations for this result. First, it is possible that there are lessened effects of on-line rhythmic compensation in fast, conversational speech, as argued by Tilsen (2002; see also Hayes 1995; a.o.). Such an explanation for the lack of a ganging-up effect of stress clashes is supported by the fact that we do actually see one-sided effects in the results: that is, when a speaker encounters a stress clash in the speech stream, phonetic adjustment for non-optimal rhythm is a highly localized phenomenon that fails to take into account what came before or what might come after. Bell et al.’s (2009:106) proposal of a casual speech production module that syncs planning and articulation predicts that such localized, myopic effects might occur when speech planning is just slow enough to allow articulatory adaptation but at the same time is just fast enough to only allow for short-term coordination. In this way, the rhythm results here line up with the behavior of word categories with respect to predictability and frequency effects, which Bell et al. (2009) use their production model to explain. A second explanation for the lack of a significant interaction of surrounding stress clash environments is that we are only looking here at a narrow definition of rhythm embodied as stress clash. Effects from longer time-course rhythmic planning (e.g., lapse, triple rhythms, footing, etc.) are not captured here in the model, and these effects could affect the evaluation of rhythm.

The main take-away point from these results is that some word clusters—namely, the ones that primarily contain function words and exhibit greater reducibility potential—are the ones that will utilize more phonetic adjustments when a non-optimal rhythmic environment is encountered. Words in Clusters 1 and 2 pattern together in having no significant differences in their rhythm-based behaviors, and they only demonstrate sensitivity to stress clash in duration reduction. The fact that only Cluster 3 and 4 words demonstrate significant effects in the segmental reduction condition is telling. Because segmental reduction indicates contrastive, categorical differences in phone quality, this result indicates that only clusters with greater reducibility potential resort to more drastic unfaithfulness in order to accommodate non-optimal rhythm. Clusters 1 and 2 do accommodate for non-optimal rhythm as well, but they only do so via durational reduction, which "costs" less in terms of faithfulness violation from the target pronunciation.

3 Syntactic choice, revisited

We return here to one of the main questions in this thesis, regarding the cross-modular nature and effects of rhythmic regularity: how much phonetic and phonological material is available during syntactic operations? The first portion of the current chapter has shown that word categories display differential use of phonetic adjustment for non-optimal rhythm. Following from this result, it is necessary to consider how phonetic adjustment relates to extra-phonological repairs for rhythmic irregularity. In the genitive alternation, for example, the *of* in the *of*-genitive construction is a preposition, and according to the results of Chapter 4 and §2 of this chapter, may demonstrate greater variability in phonetic pronunciation. Therefore, two alternative repairs exist for long stress lapses: a phonetically unreduced (i.e., stressed) *of*, as in (55a), or a syntactic alternative construction, as in (55b).

(55) *the fámily of the víctim* →

a. the fámily **óf** the víctim
b. the víctim's fámily

Shih et al.'s (to appear; see also Chapter 4) study of genitive construction choice only took into account the syntactic repair for stress lapse sequences. However, if a phonetic option is also available, does the speaker have forward knowledge of the phonetic option? If so, then models of language production must allow surface-stream information feedback into the syntactic computation, which drastically contradicts many of the currently accepted models of language production and grammar (e.g., Bock & Levelt 1994; Ferreira & Slevc 2007) (see discussion in Chapter 1). A related issue concerns the representation of phonetic knowledge in the mental lexicon. Given that we see differences in reduction patterns between word categories, how are these differences represented, and do they arise directly from encoded differences in lexical stress?

Two exploratory studies are presented here that address these issues and to serve as guidelines for future work. The first study (§3.1) examines the availability of phonetic adjustments in relation to extra-phonological adjustments for non-optimal rhythmic sequences. The second study (§3.2) investigates whether a four-part word category division is directly encoded as differences in levels of lexical stress. The findings presented here point to the necessity to maintain some division between phonetic information and more high-level concepts of syntax and phonological stress, but they also suggest that reducibility potential has crucial consequences for processes sensitive to lexical stress.

3.1 Phonetic adjustments in rhythmic computation

This first study revisits the genitive construction choice data and results presented in Chapter 4. A revised measure of rhythm is developed, which incorporates the option for phonetic adjustments. This measure of rhythm is tested in the genitives dataset. The results demonstrate a drop in explanatory power of rhythm when phonetic adjustments are taken into account, which suggests that, at the level when syntactic choices are made, no phonetic adjustment information is available to the speaker.

3.1.1 A new rhythmic computation

As demonstrated in (55) above, monosyllabic function words in the midst of a long lapse have the ability to remain in their unreduced forms so as to repair rhythmic alternation. Taking the example given above, the revised measure of *of*-Eurhythmy Distance (see definition in Chapter 4) would produce a value of 1 versus 3 in the original *of*-ED measure.

An automatic parser was built in Python to determine whether a word would be produced in weak or strong form phonetically, taking into account surrounding rhythmic environment and the scalar stress of word categories. Stressed and unstressed syllables of polysyllabic words were treated as invariably strong and weak, respectively. Monosyllabic Cluster 1 words were also treated as invariably strong, under the assumption that these words have very little reducibility potential. The remainder of the monosyllabic words were assigned a value in a four-part scale (shown in 56).

(56) Scalar stress values

1	Stressed syllables of polysyllabic words; Cluster 1 monosyllables
0.75	Cluster 2 monosyllables
0.5	Cluster 3 monosyllables
0.25	Cluster 4 monosyllables
0	unstressed syllables of polysyllabic words

As the parser progresses from left to right, if a target word (i.e., a monosyllabic word) has an equal or higher scalar stress value than its neighboring syllables, it is coded as strong; otherwise, it is coded as weak. For example, because *of* in *the family of the victim* is preceded by an unstressed syllable of a polysyllabic word and followed by a Cluster 3 monosyllabic word, it is scanned as strong, as illustrated in (57). The article *the* in the same example is not scanned as strong because it is immediately followed by a stressed syllable of a polysyllabic word.

(57)

the	fá	mi	ly	of	the	víc	tim
0	1	0	0	0.5	0.5	1	0
W	S	W	W	S	W	S	W

Revised *of*-ED and *s*-ED measures that incorporate these phonetic adjustments (referred to here as *of*-ED_{ph} and *s*-ED_{ph}) were then calculated using the algorithm introduced in Chapter 3.

3.1.2 Results

The revised Eurhythmy Distance measures were tested with the same control models in Shih et al. (to appear; see also Chapter 4). The results are provided in Table 3.

Table 29. Logistic regression estimates: Ratios represent the relative chances of *s*-genitive over *of*-genitive.

Factor	Estimate	Std. Error	Z value	Pr (>|z|)	
Intercept	-0.613	0.0985	-6.22	< 0.0001	***
Possessor animacy = inanimate	-3.6264	0.2036	-17.81	< 0.0001	***
Word count (log diff)	-3.2287	0.5515	-5.85	< 0.0001	***
Final sibilant	-1.0983	0.2931	-3.75	0.0002	**
Semantic relation = prototypical	0.9653	0.3041	3.17	0.0015	*
s-ED_{ph}	-0.5293	0.6326	-0.84	0.4028	
of-ED_{ph}	0.6419	0.4475	1.43	0.1514	
Possessor givenness = not given	0.4049	0.2541	1.59	0.1111	
Possessor freq (log)	-0.0109	0.2292	-0.05	0.962	
Persistence	0.3269	0.2181	1.5	0.1339	
Speaker birthdate	0.0034	0.0018	1.91	0.0563	
Speaker sex = M	-0.3816	0.1908	-2.00	0.0455	.
Interactions					
s-ED_{ph} * animacy = inanim	2.3218	1.2455	1.86	0.0623	
of-ED_{ph} * animacy = inanim	-0.449	0.8704	-0.52	0.606	
N	1107	R^2		0.653	
model χ^2	733.08	%correct (%baseline)		91.6 (69.53)	
Dxy	0.832	AIC_c		794.9644	

. significant at $p < 0.05$, * significant at $p < 0.01$, ** significant at $p < 0.001$, *** significant at $p < 0.0001$

As the results in Table 3 show, the revised Eurhythmy Distance measures play no reliable role in predicting genitive construction choice. The model also uncovers no significant interaction between rhythm and animacy, as we saw with the original Eurhythmy Distance measure. Comparing a model using ED_{ph} and the original measure of ED, we see a significant drop in the weight of evidence for a model using the revised ED_{ph} measure ($\Delta AIC_c = 15.7346$). The original ED measure also does better in terms of variance explained and predictive accuracy ($R^2 = 0.663$, $C = 0.92$ versus $R^2 = 0.653$, $C = 0.916$).

The comparison of these two measures of rhythm suggests that knowledge of phonetic adjustments downstream is not available at the time that syntactic choices for fixing non-optimal rhythm must be made (although this is not to say that phonetic adjustments can apply post hoc). For example, given the possessor-possessum pair, *today* and *technology* (taken from the dataset), animacy predisposes this pair for an *of*-genitive realization because the possessor *today* is inanimate.[64] Under the original calculation of *of*-ED, the *of*-genitive alternative is heavily penalized as demonstrating a long lapse:

(58) *technólogy of todáy*
W S WWWW S = 4 – 1 = 3 *of*-ED

The revised calculation of *of*-ED_{ph}, however, takes into account the possibility that *of* could be phonetically stressed on the surface in order to avoid such a long lapse, as illustrated in (59).

(59) *technólogy of todáy*
0 1 0 0 0.5 0 1
W S WW S W S = 2 – 1 = 1 *of*-ED

For *of*-ED_{ph}, then, the *of*-genitive alternative for *technology of today* is not too rhythmically offensive because the rhythmic irregularity can be repaired via phonetic adjustment. In such cases, the *of*-ED measure is more likely to predict that the *s*-genitive

[64] Rosenbach (2005) points out that temporal genitives of this type tend to prefer possessor-first order: e.g., *today's technology*; *yesterday's news*. The size of the genitives dataset will have to be expanded before we can tease out such fine-grained inferences between phonological and semantic factors. However, it should be noted that it is not necessarily the case that these factors are always independent—or independently used by speakers. For example, it has been shown that phonological, syntactic, and semantic factors often align (e.g., function words are lighter in phonological material as well as semantic content), and speakers could be using all available information to strengthen cues. Therefore, just because there may be semantic factors does not mean that rhythmic factors that align with semantic cues do not also contribute: Hanssen et al. (2013), for instance, has recently shown that rhythmic patterns in word order alternatives in fact affect semantic interpretations.

alternative will be chosen in order to avoid lapse, whereas the *of*-ED_{ph} measure does not penalize such structures as heavily. Herein lies the weakness of *of*-ED_{ph} in model prediction because such instances do in fact prefer to utilize the *s*-genitive alternative: for example, the possessor-possessum pair given above appears in the dataset as *today's technology*

It is evident that a certain amount of phonological information must be available at the moment of syntactic choice. There is the effect of rhythm, as demonstrated in Chapter 3, and there is the strong effect of adjacent sibilant avoidance for genitive construction choice. It is furthermore evident that the type of phonological information that must be available cannot be limited to high-level prosodic information. In order for rhythm and OCP-type preferences to affect construction choice, low-level phonology—that is, metrical structure and segmental material—must be available. However, the results presented in this section suggest that some limits may yet still be necessary for the interaction between sound and structure. Namely, the availability of a phonetic adjustment and the mapping of reduced versus strong forms of monosyllabic words does not seem to play into rhythmic assessment when extra-phonological adjustments such as construction choice are available.

3.2 Scalar stress in weight computation

The assessment of rhythm raises the critical issue of what the lexical representation of the four word categories uncovered in Chapter 5 (and §2 of the current chapter) is. This is a critical issue that has been hitherto glossed over and will only be addressed cursorily here. The results in §3.1 above already suggest that some privileged, abstract form of stress representation must be at work in the computation of rhythm—that is, we find that it is not scalar stress but a binary alternation between strong and weak syllables that is predictive as a measure of rhythm for genitive construction choice.

Several proposals have been put forth for representing the dichotomy between content and function words in the lexicon. One common approach is to assume separate lexicons for word classes—or, equivalently, that words are diacritically marked for *Lex* or *Fnc* status in the syntax (that gets passed to the phonology). Under this view of the lexicon, the reducibility potential of the four categories of content and function words could be indexed to each separate lexicon or diacritic. A competing conceptualization for the difference between content and function words is to directly encode word types as the presence or absence of underlying lexical stress (e.g., Inkelas and Zec 1993). Under this latter view, the four-part categorization of content and function words proposed here would have to be represented as a four differing levels of lexical stress (see also Chomsky and Halle 1968 for a representation of stress as four-leveled). These approaches are largely equivalent in terms of predicting function and content word reduction, but a crucial difference arises when one considers the consequences for stress-sensitive operations—for example, rhythm, as discussed above. If the four word categories are built in as four levels of underlying lexical stress, then we would expect stress-sensitive operations to be able to reference these levels.

One testing ground for stress-dependent phenomena is that of end weight effects, as introduced briefly in Chapter 4. It has been observed that heavier and longer constituents tend to come at the periphery of phrases; in English, this means that ordering prefers heavier constituents at the end—for example, *peas and carrots* versus *carrots and peas*. In terms of syntactic construction choice, numerous measures of weight have been proposed, ranging from processing-based to syntactic to phonological (Zec & Inkelas 1990; Hawkins 1994; Zubizarreta 1998; Gibson 2000; Wasow 2002; Temperley 2006; Anttila et al. 2010; see Grafmiller & Shih 2011 for an overview of weight measures for the English genitive alternation). In a study of weight measures in the English genitive alternation, Grafmiller & Shih (2011) concluded that the number of primary stresses was one of the most important and reliable measures of heaviness for the purposes of end weight. This conclusion follows from theories of end weight based on the attraction of stress to prosodically strong positions at the ends of phrases (Zubizarreta 1998; Anttila et al. 2010).

If stress is attracted to prosodically strong positions in phrasal domains, then a scalar stress representation underlyingly for different word categories should be reflected in end weight. For example, for the possessor-possessum pair *our court system* and *the authority* in (60), a measure of primary stresses would indicate that the possessor is only one point longer than the possessum (assuming scalar stress values provided in (56)). A measure of scalar stress, however, would find the possessor to be 1.5 points longer than the possessum, given that *our* is a Cluster 3 word.

(60) the authórity of our cóurt sýstem
x 0 1 00 x 0.5 1 1 0

As a consequence, we would expect, under a scalar stress-based measure of weight, that heaviness could be induced by certain function words as well. This section presents a replication of Grafmiller & Shih's (2011) study to compare the effect of measuring weight using a four-part lexical stress representation. The findings demonstrate that a scalar notion of stress contributes important additional information to the determination of end weight beyond primary stress. As such, these results provide preliminary evidence that word category differences may have direct encoding in underlying lexical stress representations.

3.2.1 Data and methodology

This study uses the genitive data presented in Chapter 4 and replicates the methodology from Grafmiller & Shih (2011). In their study comparing weight measures, Grafmiller & Shih test multiple hypotheses on how to measure weight in both genitive and dative alternation in American English: syntactic complexity, dependency length, primary stress weight, weight by number of syllables, weight by number of orthographical words (see also Szmrecsányi 2004 for a similar comparison, except without phonological predictors). The current study utilizes all of the weight measures that Grafmiller & Shih tested, with the exception of the number of syllables. The number of syllables as a measure of weight was found by Grafmiller & Shih to be across-the-board not a robust measure of weight, so it is left out here; however, it is possible that the number of syllables (or the phonological complexity within syllables and segments) is crucial to the measure of weight if the domain of weight computation is more limited (e.g., Ingason & MacKenzie 2011; Ryan 2013).

Comparisons between weight measures are carried out via residualization, conditional random forest analysis and model AIC comparisons (see Chapters 4 and 5, respectively) in order to alleviate collinearity difficulties in assessing overlapping weight measures. For additional overviews of the methodologies, see Grafmiller & Shih 2011.

3.2.2 Results and discussion

We first examine whether a four-part scalar stress representation of stress contributes any relevant and independent information beyond primary stress to a measure of end weight. Results of a model with the residual information of scalar stress after primary stress has been taken into account is given in Table 30.

Table 30. Logistic regression estimates (abridged)

Factor	Estimate	Std. Error	Z value	Pr (>|z|)	
Intercept	-0.6904	0.1025	-6.73	< 0.0001	***
Primary stresses (log diff)	-2.2038	0.5094	-4.33	< 0.0001	***
Scalar stress (log diff) residual	-3.7646	2.0639	-1.82	0.0682	
...					
N	1107	R^2		0.654	
model χ^2	734.43	%correct (%baseline)		91.6 (69.53)	
Dxy	0.831	AIC_c		794.671	

. significant at $p < 0.05$, * significant at $p < 0.01$, ** significant at $p < 0.001$, *** significant at $p < 0.0001$

The results of a residualized scalar stress measure trend towards significance, although it's not quite there—perhaps due to the limited size of the dataset. We see that the effect does trend in the expected direction: as the possessor gets heavier than the possessum (i.e., has more stress), the choice of an *of*-genitive construction becomes more likely. This result suggests that scalar stress may offer critical information in the computation of heaviness for end weight effects that primary stress alone does not.

We can compare the weight of evidence for a heaviness measure that uses scalar stress information against the other measures of end weight via AIC_c comparisons, given in Table 31.

Table 31. Model AIC_c

Weight measure	AIC_c	Δ_i
Syntactic nodes	764.876	0
Referents	781.6412	16.7652
Scalar, four-level stress	793.2502	28.3742
Orthographic words	794.9644	30.0884
Primary stresses	795.9663	31.0903

As Grafmiller & Shih (2011) show, syntactic nodes are one of the best individual measures of end weight for genitive construction choice. Amongst the phonological measures, we see that the addition of a scalar, four-level stress measure of end weight actually does well as an individual predictor of the genitive alternation, as compared to primary stresses and orthographic words. Its presence in the model set demotes the ranking of a model that uses primary stresses as a measure of end weight. It should be noted, however, that the differences between the bottom three models in Table 31 are extremely small, which is expected to a certain extent because of the high degree of overlap and correlation between these three measures of end weight.

Finally, we examine the results of a random forest model of genitive construction choice that tests the independent contributions of end weight measures in a way that is not tied to hidden dependencies between predictors (see discussion in §3.2.1

and in Chapter 4). The results are provided in Table 32,[65] and they parallel the AIC$_c$ model comparison results discussed above.[66]

Table 32. Random forest variable importance

Weight measure	Variable importance
Syntactic nodes	0.002958435

[65] Referent count was temporarily excluded from this analysis due to computational technical difficulties, which may or may not be resolvable.

[66] The model comparison methodology also comes with its own version of variable importance, as presented in Burnham & Anderson (2002; 2004:274). Model comparison-based variable importance is calculated as the cumulative probability of a given predictor in models of the entire model candidate space that display the highest Akaike weight values (i.e., models that are most likely to be better approximations of reality given the evidence available). Grafmiller & Shih (2011) utilize this approach for an additional view into testing weight measures, and find mostly similar results between model comparison variable importance and random forest variable importance.

Utilizing model comparison variable importance here, however, produces results that run counter not only to the model AIC$_c$ rankings in Table 31 but also the random forest results in Table 32. Model comparison-based variable importance rankings are given below. These results were generated using the `dredge()` and `importance()` functions of `R` package `MuMIn` (Bartoń 2013).

Variable importance (calculated from AIC$_c$ weights)	
Weight measure	**Variable importance**
Referents	0.9995325
Primary stresses	0.9937449
Syntactic nodes	0.9833057
Orthographic words	0.8638455
Scalar, four-level stress	0.4086998
Syllables	0.3449252

As can be seen from the model comparison variable importance results, a primary stress-based measure of end weight ranks higher than a scalar, four-level stress measure. One caveat to note here is that AIC$_c$-based variable importance methods are not completely immune from the harmful effects of collinearity in inference because Akaike weight estimation still depends on the performance of whole models containing collinear variables; rankings may reflect unseen dependencies within the data between variables. Random forests are a more reliable—though computationally less convenient—method of untangling individual contributions of highly collinear predictors.

Scalar, four-level stress	0.0008264059
Orthographic words	0.0006894866
Primary stresses	0.0002640587

These results are by no means deterministic given the size of the dataset, but we can take them as preliminary evidence suggesting that lexical stress—at least in the computation of end weight, if we buy into some stress-based theory of end weight—should encode a scalar notion of word categories. Future research will be necessary to confirm or revise these findings.

4 Conclusion

The results presented in this chapter lead to a view of the linguistic system in which some separation between grammatical informations remains an integral feature, despite the necessity of allowing phonology to condition morphosyntactic phenomena. The first case study demonstrated that lexical word categories display different behavior in terms of phonetic adjustments in the optimization of rhythmic patterning, which suggests that the type of compensation strategy that speakers utilize depends on lexical word categories involved at varying stages of access.

The results of the second case study reveal that the output of phonetic computations of rhythm—for instance, the post-lexical stressing or reduction of function words in rhythmic optimization—is unavailable at the point of morphosyntactic choice. This means that phonetic, surface optimization cannot compete with morphosyntactic optimization of rhythm. It still remains to be see whether the output of *phonological* (rather than phonetic) processes is available to block morphosyntactic optimization. The contribution here is merely to dispel the null hypothesis that all information is equally available at the point of morphosyntactic computation.

The conclusion of the final case study offers insight into the type of phonological information that is available at the interface with morphosyntax: namely, underlying phonological properties. It was shown that the four-category stress properties of lexical word categories presented in Chapter 5 may play a role in determining the weight of constituents in heavy-last effects of genitive construction choice. Such a result suggests that lexically-encoded, underlying phonological information *before* the output of surface, post-lexical processes must be available at the stage of morphosyntactic computation. Under this view, we would not expect phonetic duration, for instance, to be as good of a predictor of weight; investigation of this issue is saved for future study.

Chapter 7. Conclusion

The overarching goal of this thesis was to examine the optimization of phonological well-formedness via morphosyntactic choices. Rhythmic alternation offers a natural window into probing the interaction between phonology and morphosyntactic constituents because of its syntagmatic properties, operating across morpheme and word boundaries. The key question at hand in this work was how morphosyntactic processes are influenced (if at all) by a phonological drive towards optimal rhythm.

It was argued in Chapter 2 that phonologically-conditioned syntactic phenomena are of the same family as more familiar phonologically-conditioned morphological phenomena. Both sets of phenomena share similar phonological conditions that can trigger a set of defined responses in the face of phonological pressures: blocking, suppletion, periphrasis, ordering, and deletion. While phonologically-conditioned syntactic effects may be weaker than morphological ones, we naturally expect them to be so, given [1] the generally-observed decay of phonological constraint strength across constituent boundaries and [2] the increasing number of extra-phonological (i.e., syntactic, semantic, usage-based) factors that phonological constraints must compete with in the assessment of morphosyntactic competitors. Furthermore, the phonological conditions that appear to affect morphological and syntactic material are ones that regulate syntagmatic configurations (e.g., phonotactics, rhythm). Constraints that regulate non-contextual effects—for instance *[labial] from Zwicky & Pullum's famous straw man example of phonologically-conditioned syntax—will not be commonly violated by morphosyntactic operations and therefore, their effects in syntax will be extremely limited.[67] A contextual constraint of the type *[labial][labial] (common in approaches

[67] For example, *[labial] might cause word choice differences, but it is unlikely to cause word order or other structural alternations that would not repair a *[labial] violation. The only syntactic repair of such a constraint would be to choose a different syntactic construction altogether that avoids the offending labial-initial word. In this case, you would not witness any overt difference in the linearization of

to phonotactics: see e.g., Hayes & Wilson 2008), on the other hand, would be violated by morphosyntactic combinatoric processes.

Chapter 3 turned to an example of a type of suppletive process that is conditioned by phonological factors: the replacement of one lexeme with another in name choice in English. In a study of forename and surname pairs, it was shown that the choice of forename to match a given surname is influenced by a number of phonological constraints: rhythmic well-formedness, OCP, and similarity. Although name pairs are a special, often carefully-considered subset of linguistic phenomena, we expect that the intuitions of euphony for a chosen name pair should arise from natural intuitions of phonological well-formedness that are at work in English. Moreover, name pairs provided a syntagmatic construction relatively free from compounding syntactic and semantic factors that frequently overshadow phonological effects in syntactic phenomena.

Chapter 4 presented an analysis of a syntactic alternation between two genitive constructions in English: the cliticized *–'s* form (*car's wheel*) versus the more "periphrastic" *of* form (*the wheel of the car*). Previous work has already shown that phonological factors such as OCP sibilant avoidance is at work in genitive construction choice. This chapter showed further that genitive construction choice also, to a certain extent, optimizes for rhythmic well-formedness. The role of rhythm is complex, however, and involves interactions with animacy.

Given the growing evidence in recent work and the first half of this thesis demonstrating phonologically-conditioned syntactic phenomena, the issue becomes to what extent can low-level information in not only phonology but also the surface speech stream affect (morpho-)syntactic encoding? Chapters 5 and 6 took up this issue via a case study of content and function word stress in the phonological conditioning of morphosyntactic choices. As a preliminary step, Chapter 5 presented a study of the surface, phonetic differences in reduction between words. It was found that words fall on a four-part cline of stress and reducibility that largely but not completely correlates

words; furthermore, doing so might be overly costly in the system, thus offering a partial explanation towards the rarity (and implausibility) of such phonologically-conditioned syntactic effects.

with traditional assumptions of content and function word categories. This division crucially demonstrates the complexity of phonological versus phonetic encoding of stress information, which affects how underlying versus surface rhythmic patterns influence morphosyntactic computation. Chapter 6 demonstrated that morphosyntactic computation is insensitive to the consideration of surface rhythmic patterns but remains sensitive to underlying phonological stress information.

In summary, the main contribution of this thesis is an empirical one. An early assumption in linguistic theory of Phonology-free Syntax has—irrespective of theoretical implementation—biased the field against the possibility of phonologically-conditioned syntactic effects. The empirical conclusions presented in this thesis are meant to help dispel this strong view of the role of phonology across morpheme and word boundaries. These conclusions do not necessarily mean that phonology exists in the syntactic module architecturally, which is too complex of an issue to be taken up here. The simple point instead is that these effects of phonologically-conditioned syntax, which are [1] not as uncommon as previously assumed, [2] more robust than previously assumed, and [3] parallel familiar morphological phenomena, should be accounted for in the design of the phonology-morphosyntax architecture.

Appendix A. Random forest variable importance for genitives model

Table 33. Random forest variable importance

Weight measure	Variable importance
Possessor animacy	0.176731
Word count (log diff)	0.006317848
of-Eurhythmy Distance	0.001046455
Persistence	0.0005867971
Semantic role	0.0003129584
Final sibilant	0.0001858191
Possessor log frequency	0.00000801956
s-Eurhythmy Distance	-0.0000391198

Appendix B. Alternative cluster models

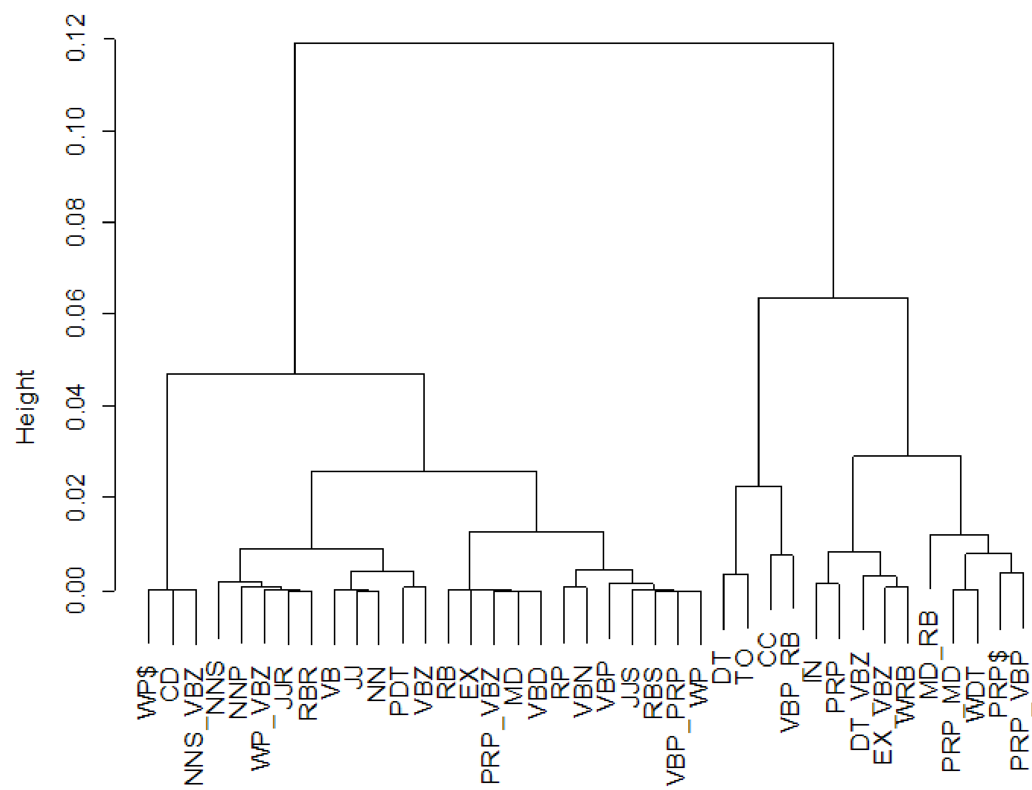

Figure 29. Complete linkage hierarchical cluster dendrogram: Part-of-speech clusters by phone distance and duration distance (cf. Ward's method hierarchical cluster in Chapter 5).

Appendix C. Closed class versus open class verbs

Verb form	Closed				Open
Base			*have*	*do*	Lexical
VB			612	893	8237
Past tense	*was*	*were*	*had*	*did*	Lexical
VBD	3018	740	965	435	4714
Present tense, not 3ps.	*am*	*are*	*have*	*do*	Lexical
VBP	65	1075	1477	697	7232
Present tense, 3ps.	*is*		*has*	*does*	Lexical
VBZ	1866		365	164	1440
Past participle	*been*		*had*	*done*	Lexical
VBN	556		10	182	1406

Appendix D. Clusters by Penn Treebank part-of-speech tags

1. Standard content versus function division

Content vs. Function	Parts-of-speech
Content	cardinal number, possessive *wh* pronoun, noun_verb
	adjective, comparative adjective, noun (singular and plural), proper noun, predeterminer, comparative adverb, verb, present verb (3rd ps.), *wh* pronoun_verb (pres., 3rd ps.)
	existential *there*, superlative adjective, modal auxiliary, adverb, superlative adverb, particle, past participle verb, past tense verb, present verb (not 3rd ps.), *wh* pronoun, pronoun_verb (pres., 3rd ps.), verb_pronoun
Function	preposition, pronoun, pronoun (possessive), *wh* determiner, *wh* adverb, determiner_verb (pres.), *there*_verb (pres.), modal_adverb, pronoun_modal, pronoun_verb (pres., not 3rd ps.)
	coordinating conjunction, determiner, preposition *to*, verb_adverb

2. H-Clust A: 2 cluster, 4 cluster, 5 cluster

2-cluster	4-cluster	5-cluster	Parts-of-speech
Cluster 1	Cluster 1	Cluster 1	cardinal number, possessive *wh* pronoun, noun_verb
		Cluster 2	adjective, comparative adjective, noun (singular and plural), proper noun, predeterminer, comparative adverb, verb, present verb (3rd ps.), *wh* pronoun_verb (pres., 3rd ps.)
	Cluster 2	Cluster 3	existential *there*, superlative adjective, modal auxiliary, adverb, superlative adverb, particle, past participle verb, past tense verb, present verb (not 3rd ps.), *wh* pronoun, pronoun_verb (pres., 3rd ps.), verb_pronoun
Cluster 2	Cluster 3	Cluster 4	preposition, pronoun, possessive pronoun, *wh* determiner, *wh* adverb, determiner_verb (pres.), *there*_verb (pres.), modal_adverb, pro-

			noun_modal, pronoun_verb (pres., not 3rd ps.)
	Cluster 4	Cluster 5	coordinating conjunction, determiner, preposition *to*, verb_adverb

3. H-Clust B: 2 cluster, 3 cluster, 4 cluster, 5 cluster, 6 cluster, 7 cluster

2-cl	3-cl	4-cl	5-cl	6-cl	7-cl	Parts-of-speech
Cl 1	Cl 1	Cl 1	Cl 1	Cl 1	Cl 1	cardinal number, possessive *wh* pronoun, noun_verb
		Cl 2	Cl 2	Cl 2	Cl 2	adjective, comparative adjective, noun (singular and plural), proper noun, comparative adverb, lexical verb, past participles *done*, *had*, *been*, *wh* pronoun_verb (pres., 3rd ps.)
				Cl 3	Cl 3	superlative adjective, superlative adverb, predeterminer, particle, past tense lexical verb, past participle lexical verb, present lexical verb, *wh* pronoun, present verbs *is*, *has*, *does*, verb_pronoun
Cl 2	Cl 2	Cl 3	Cl 3	Cl 4	Cl 4	existential *there*, modal auxiliary, adverb, verbs *do*, *have*, past tense verbs *had*, *were*, *was*, *did*, pronoun_verb (pres, 3rd ps.)
					Cl 5	preposition, pronoun, *wh*-adverb, determiner_verb (pres., 3rd ps.), *there*_verb (pres., 3rd ps.)
	Cl 3	Cl 4	Cl 4	Cl 5	Cl 6	possessive pronoun, *wh* determiner, present verbs *are*, *have*, *do*, *am*, modal_adverb, pronoun_modal, pronoun_present verb (not 3rd ps.)
			Cl 5	Cl 6	Cl 7	coordinating conjunction, determiner, preposition *to*, present verb_adverb

4. 10-point scale (Altenberg 1987); translated to Penn Treebank part-of-speech tags

Prosodic potential	Parts-of-speech
1	common noun, lexical verb (base), past participle lexical verb, adjective, comparative adjective, superlative adjective, adverb, comparative adverb, superlative adverb, *wh* adverb
2	proper noun, past tense lexical verb, particle, predeterminer, cardinal number
3	do
4	have, been, had, done
5	(n/a for Penn Treebank part-of-speech tags)
6	*wh* pronoun, modal auxiliary, *and*, preposition, *wh* determiner
7	but
8	personal pronoun, *be, was*, *were*, *had*, *did*,
9	possessive pronoun, existential *there*
10	determiner

5. Hybrid model: content versus 5-cluster function word solution

C vs. F	Cluster	Parts-of-speech
Content		adjective, comparative adjective, noun (singular and plural), proper noun, comparative adverb, lexical verb, superlative adjective, superlative adverb, past tense lexical verb, past participle lexical verb, present lexical verb, adverb
Function	Cl 1	cardinal number, possessive *wh* pronoun, noun_verb
	Cl 2	past participles *done, had, been*, *wh* pronoun_verb (pres., 3rd ps.), predeterminer, particle, *wh* pronoun, present verbs *is*, *has*, *does*, verb_pronoun
	Cl 3	existential *there*, modal auxiliary, verbs *do, have*, past tense verbs *had*, *were*, *was*, *did*, preposition, pronoun, *wh*-adverb, pronoun_verb (pres, 3rd ps.), determiner_verb (pres., 3rd ps.), *there*_verb (pres., 3rd ps.)
	Cl 4	possessive pronoun, *wh* determiner, present verbs *are*, *have*, *do*, *am*, modal_adverb, pronoun_modal, pronoun_present verb (not 3rd ps.)
	Cl 5	coordinating conjunction, determiner, preposition *to*, present verb_adverb

Appendix E. Hybrid cluster regression model for phone distance

Table 34. Hybrid cluster regression model for phone distance

Factor	Est	Std. Err	t value	pmcmc	Pr (>|t|)
Intercept	-0.008	0.00123	-6.451	0.0001	0.0000
log(Speech rate)	0.0027	0.00039	6.841	0.0001	0.0000
log(Word frequency)	0.0014	0.00061	2.263	0.0128	0.0236
log(Bigram previous)	0.0013	0.00009	14.506	0.0001	0.0000
log(Bigram next)	0.001	0.00009	11.514	0.0001	0.0000
Accent ratio	-0.169	0.02714	-6.228	0.0001	0.0000
Phrase position = nonfinal	0.0056	0.00044	12.598	0.0001	0.0000
Coda = Y	0.0154	0.00113	13.624	0.0001	0.0000
Complex coda = Y	0.0027	0.00038	7.105	0.0001	0.0000
Onset = Y	-0.007	0.00147	-4.842	0.0001	0.0000
Word cluster = Fx 1	-0.003	0.00228	-1.207	0.1758	0.2275
= Fx 2	-0.0007	0.00087	-0.853	0.5034	0.3938
= Fx 3	0.0021	0.00042	4.907	0.0001	0.0000
= Fx 4	0.0018	0.0003	5.953	0.0001	0.0000
= Fx 5	-0.0002	0.00026	-0.907	0.4744	0.3642
Interactions					
log(Freq) * log(Bigram prev)	0.0002	0.00003	5.948	0.0001	0.0000
log(Freq) * log(Bigram next)	0.0003	0.00003	11.428	0.0001	0.0000
log(Freq) * Cluster = Fx 1	0.0001	0.0012	0.118	0.9222	0.9058
= Fx 2	-0.0013	0.00062	-2.133	0.0452	0.0329
= Fx 3	-0.00008	0.00027	-0.28	0.9236	0.7794
= Fx 4	0.00006	0.00019	0.312	0.839	0.7547
= Fx 5	-0.0007	0.00018	-3.92	0.0001	0.0001
log(Bigr prev) * Cluster=Fx1	-0.00004	0.0002	-0.214	0.8435	0.8305
= Fx 2	0.00001	0.0001	0.020	0.9888	0.9842
= Fx 3	0.0001	0.00005	2.104	0.0332	0.0354
= Fx 4	0.00023	0.00005	4.692	0.0001	0.0000
= Fx 5	0.00025	0.00003	8.078	0.0001	0.0000
log(Bigr next) * Cluster=Fx1	-0.00037	0.00019	-2.004	0.0394	0.0451
= Fx 2	0.00007	0.00011	0.622	0.5454	0.5337
= Fx 3	0.00016	0.00005	3.289	0.0016	0.001
= Fx 4	-0.00004	0.00004	-0.891	0.382	0.3731
= Fx 5	0.00015	0.00003	5.578	0.0001	0.0000
PhrPos = nonfin * Clust=Fx1	-0.0003	0.00101	-0.276	0.7822	0.7827
= Fx 2	-0.00006	0.0005	-0.128	0.9218	0.8979
= Fx 3	0.00108	0.0002	4.949	0.0001	0.0000
= Fx 4	0.00124	0.00026	4.809	0.0001	0.0000
= Fx 5	0.00006	0.00012	0.528	0.576	0.5973

Groups	Name	Variance	Std. Deviation

Word	(Intercept)	0.00015067	0.012275
Residual		0.00100549	0.031709
N	206858	Groups: word	2561
Log likelihood	418926	R^2	0.429386
Deviance	-838425	AIC_c	-837774.5

Appendix F. Individual subset regression results

Phone distance subset models

Table 35. Cluster 1 subset regression results

Factor	Est	Std. Err	t value	pmcmc	Pr (>|t|)
Intercept	-0.00008	0.000397	0.197	0.6704	0.8442
log(Speech rate)	0.00056	0.000342	1.645	0.1178	0.1
log(Word frequency)	0.00064	0.000138	4.663	0.0001	0.0000
log(Bigram previous)	0.00011	0.000029	3.727	0.0008	0.0002
log(Bigram next)	0.00021	0.000031	6.761	0.0001	0.0000
Accent ratio	-0.0125	0.02211	-0.563	0.3984	0.5733
Phrase position = nonfinal	0.00093	0.000165	5.625	0.0001	0.0000
Coda = Y	0.00334	0.000717	4.663	0.0001	0.0000
Complex coda = Y	0.00412	0.000337	12.212	0.0001	0.0000
Onset = Y	-0.0028	0.000877	-3.166	0.0001	0.0015
Interactions					
log(Freq) * log(Bigram prev)	0.00003	0.000015	2.01	0.042	0.0444
log(Freq) * log(Bigram next)	0.00005	0.000014	3.338	0.0012	0.0008

Groups	Name	Variance	Std. Deviation
Word	(Intercept)	0.00004393	0.0066285
Residual		0.00014298	0.0119574
N	46171	Groups: word	2133
Log likelihood	137642	R^2	0.3207093
Deviance	-275468	AIC_c	-275255.1

Table 36. Cluster 2 subset regression results

Factor	Est	Std. Err	t value	pmcmc	Pr (>|t|)
Intercept	-0.0013	0.00096	-1.339	0.1384	0.1804
log(Speech rate)	0.00346	0.000526	6.575	0.0001	0.0000
log(Word frequency)	0.00282	0.000489	5.776	0.0001	0.0000
log(Bigram previous)	0.00088	0.000087	10.018	0.0001	0.0000
log(Bigram next)	0.00073	0.000097	7.567	0.0001	0.0000
Accent ratio	-0.1	0.0382	-2.619	0.0018	0.0088
Phrase position = nonfinal	0.00495	0.00031	16.07	0.0001	0.0000
Coda = Y	0.01324	0.00166	7.977	0.0001	0.0000
Complex coda = Y	0.00487	0.0007	6.973	0.0001	0.0000
Onset = Y	-0.0019	0.00222	-0.847	0.3264	0.397

Interactions					
log(Freq) * log(Bigram prev)	0.00031	0.000077	4.072	0.0001	0.0000
log(Freq) * log(Bigram next)	0.00044	0.000078	5.617	0.0001	0.0000
Groups	Name	Variance		Std. Deviation	
Word	(Intercept)	0.000073		0.0085437	
Residual		0.000505		0.0224683	
N	47732	Groups: word		640	
Log likelihood	113063	R^2		0.230326	
Deviance	-226288	AIC_c		-226098.9	

Table 37. Cluster 3 subset regression results

Factor	Est	Std. Err	t value	pmcmc	Pr (>|t|)
Intercept	-0.0026	0.004365	-0.594	0.583	0.5522
log(Speech rate)	0.00046	0.000722	0.633	0.5144	0.5264
log(Word frequency)	0.00532	0.002838	1.875	0.034	0.0607
log(Bigram previous)	0.0015	0.000128	11.725	0.0001	0.0000
log(Bigram next)	0.00185	0.000122	15.169	0.0001	0.0000
Accent ratio	-0.4526	0.2621	-1.727	0.0392	0.0842
Phrase position = nonfinal	0.00749	0.000497	15.065	0.0001	0.0000
Coda = Y	0.04395	0.003246	13.542	0.0001	0.0000
Complex coda = Y	0.00025	0.000569	0.432	0.6934	0.6655
Onset = Y	0.00462	0.005846	0.791	0.4514	0.4291
Interactions					
log(Freq) * log(Bigram prev)	0.00008	0.000169	0.452	0.6782	0.6514
log(Freq) * log(Bigram next)	0.0019	0.000146	13.028	0.0001	0.0000
Groups	Name	Variance		Std. Deviation	
Word	(Intercept)	0.00061031		0.024704	
Residual		0.00122369		0.034981	
N	67295	Groups: word		94	
Log likelihood	129891	R^2		0.2026673	
Deviance	-259924	AIC_c		-259753.9	

Table 38. Cluster 4 subset regression results

Factor	Est	Std. Err	t value	pmcmc	Pr (>|t|)
Intercept	-0.00835	0.008311	-1.005	0.2506	0.315
log(Speech rate)	0.0064	0.001146	5.583	0.0001	0.0000
log(Word frequency)	0.02166	0.008269	2.619	0.003	0.0088
log(Bigram previous)	0.00421	0.000237	17.822	0.0001	0.0000

log(Bigram next)	0.00327	0.000184	17.807	0.0001	0.0000
Accent ratio	-0.4677	0.396257	-1.18	0.1598	0.2379
Phrase position = nonfinal	0.00638	0.000584	10.917	0.0001	0.0000
Coda = Y	0.03004	0.005491	5.471	0.0002	0.0000
Complex coda = Y	0.00301	0.001589	1.891	0.0418	0.0586
Onset = Y	-0.02058	0.012042	-1.709	0.0398	0.0875
Interactions					
log(Freq) * log(Bigram prev)	0.00208	0.000437	4.753	0.0002	0.0000
log(Freq) * log(Bigram next)	0.0021	0.000311	6.762	0.0001	0.0000

Groups	Name	Variance	Std. Deviation
Word	(Intercept)	0.00069541	0.026371
Residual		0.00151974	0.038984
N	45660	Groups: word	24
Log likelihood	83239	R^2	0.3516412
Deviance	-166606	AIC_c	-166450.8

Duration distance subset models

Table 39. Cluster 1 subset regression results

Factor	Est	Std. Err	t value
Intercept	0.00009	0.000065	1.31
log(Speech rate)	0.00199	0.000058	33.99
log(Word frequency)	0.00017	0.000023	7.43
log(Bigram previous)	0.00009	0.000005	17.62
log(Bigram next)	0.00011	0.000005	21.15
Accent ratio	-0.00089	0.003647	-0.24
Phrase position = nonfinal	0.00232	0.000028	82.81
Coda = Y	0.00116	0.000119	9.72
Complex coda = Y	0.00051	0.000567	8.99
Onset = Y	0.00121	0.000146	8.28
Interactions			
log(Freq) * log(Bigram prev)	0.00001	0.000002	4.56
log(Freq) * log(Bigram next)	0.00002	0.000002	10.53

Groups	Name	Variance	Std. Deviation
Word	(Intercept)	0.0000011826	0.0010875
Residual		0.0000041683	0.0020416
N	46171	Groups: word	2133
Log likelihood	219276	R^2	0.3856037
Deviance	-438780	AIC_c	-438524.2

Table 40. Cluster 2 subset regression results

Factor	Est	Std. Err	t value
Intercept	-0.00008	0.000083	-0.95
log(Speech rate)	0.0014	0.000044	31.83
log(Word frequency)	0.00002	0.000042	4.73
log(Bigram previous)	0.00001	0.000007	17.94
log(Bigram next)	0.00016	0.000008	20.22
Accent ratio	-0.00356	0.003303	-1.08
Phrase position = nonfinal	0.00235	0.000026	91.14
Coda = Y	0.00046	0.000142	3.23
Complex coda = Y	0.00014	0.000059	2.41
Onset = Y	0.00028	0.000191	1.45
Interactions			
log(Freq) * log(Bigram prev)	0.00004	0.000006	5.72
log(Freq) * log(Bigram next)	0.00005	0.000007	8.41

Groups	Name	Variance	Std. Deviation
Word	(Intercept)	0.0000006	0.00074557
Residual		0.0000035	0.0018802
N	47732	Groups: word	640
Log likelihood	231432	R^2	0.3623706
Deviance	-463086	AIC_c	-462836.8

Table 41. Cluster 3 subset regression results

Factor	Est	Std. Err	t value
Intercept	-0.00033	0.00026	1.25
log(Speech rate)	0.00221	0.00009	23.52
log(Word frequency)	0.00006	0.00019	0.33
log(Bigram previous)	0.00031	0.00002	18.77
log(Bigram next)	0.00043	0.00002	27.17
Accent ratio	-0.0468	0.01596	-2.93
Phrase position = nonfinal	0.00358	0.000065	55.23
Coda = Y	-0.0006	0.000312	-1.8
Complex coda = Y	-0.0002	0.000073	-2.12
Onset = Y	0.00056	0.000371	1.49
Interactions			
log(Freq) * log(Bigram prev)	0.00007	0.000022	3.24
log(Freq) * log(Bigram next)	0.00033	0.000019	17.45

Groups	Name	Variance	Std. Deviation
Word	(Intercept)	0.0000019562	0.0013986

Residual		0.000020788	0.0045594
N	67295	Groups: word	94
Log likelihood	267049	R^2	0.1359145
Deviance	-534296	AIC_c	-534070.7

Table 42. Cluster 4 subset regression results

Factor	Est	Std. Err	t value
Intercept	0.00063	0.000593	1.06
log(Speech rate)	0.00211	0.000092	22.98
log(Word frequency)	-0.00012	0.000602	-0.20
log(Bigram previous)	0.00037	0.000019	19.64
log(Bigram next)	0.00054	0.000015	36.88
Accent ratio	-0.0298	0.02842	-1.05
Phrase position = nonfinal	0.00321	0.000047	68.48
Coda = Y	-0.00235	0.000434	-5.43
Complex coda = Y	0.00008	0.000127	0.63
Onset = Y	-0.00136	0.00086	-1.58
Interactions			
log(Freq) * log(Bigram prev)	0.00019	0.000035	5.47
log(Freq) * log(Bigram next)	0.00031	0.000025	12.63

Groups	Name	Variance	Std. Deviation
Word	(Intercept)	0.0000035185	0.0018758
Residual		0.0000097562	0.0031235
N	45660	Groups: word	24
Log likelihood	198466	R^2	0.29771
Deviance	-397121	AIC_c	-396904.2

References

Abercrombie, David. 1967. *Elements of General Phonetics*. Edinburgh, UK: Edinburgh University Press.

Abondolo, Daniel. 1988. *Hungarian Inflectional Morphology*. Budapest: Akadémiai Kiadó.

Adams, Matthew Edward. to appear. Prosodic well-formedness and comparative grammaticality: Morphology and periphrasis in the English comparatve. In Vera Gribanova & Stephanie S Shih (eds.), *The morphosyntax-phonology connection: locality and directionality at the interface*. Oxford, UK: Oxford University Press.

Adams, Matthew Edward. 2014. The comparative grammaticality of the English comparative. Stanford, CA: Stanford University Ph.D. dissertation.

Agbayani, Brian & Chris Golston. 2010. Phonological movement in Classical Greek. *Language* 86(1). 133–167.

Agbayani, Brian, Chris Golston & Dasha Henderer. 2011. Prosodic movement. In Mary Byram Washburn & et al. (eds.), *Proceedings of the 28th West Coast Conference on Formal Linguistics*, 231–239. Somerville, MA: Cascadilla.

Akaike, Hirotugu. 1973. Information Theory as an Extension of the Maximum Likelihood Principle. In B.N. Petrov & F. Csaki (eds.), *Second International Symposium on Information Theory*, 267–281. Budapest: Akademiai Kiado.

Altenberg, Bengt. 1987. *Prosodic patterns in spoken English: Studies in the correlation between prosody and grammar for text-to-speech conversion*. (Lund Studies in English 76). Sweden: Lund University Press.

Anderson, David R. & Kenneth P. Burnham. 2002. Avoiding pitfalls when using information-theoretic methods. *Journal of Wildlife Management* 66(3). 912–918.

Anttila, Arto. 1997. Deriving variation from grammar. In Frans Hinskens, Roeland van Hout & W. Leo Wetzels (eds.), *Variation, Change, and Phonological Theory*, 35–68. Amsterdam: John Benjamins.

Anttila, Arto. 2002. Morphologically Conditioned Phonological Alternations. *Natural Language & Linguistic Theory* 20(1). 1–42.

Anttila, Arto. 2012. Stress and auxiliary reduction in English. Paper. Paper presented at the Workshop on Locality and Directionality at the Morphosyntax-Phonology Interface, Stanford, California: Stanford University. http://www.stanford.edu/~gribanov/downloads/Anttila.pdf.

Anttila, Arto, Matthew Adams & Michael Speriosu. 2010. The role of prosody in the English dative alternation. *Language and Cognitive Processes* 25(7-9). 946–981.

Arvaniti, Amalia. 2009. Rhythm, Timing and the Timing of Rhythm. *Phonetica* 66(1-2). 46–63.

Arvaniti, Amalia. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40(3). 351–373.

Auer, Peter, Elizabeth Couper-Kuhlen & Frank Müller. 1999. *Language in Time: The Rhythm and Tempo of Spoken Interaction*. Oxford, UK: Oxford University Press.

Aylett, Matthew & Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1). 31–56.

Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge, UK: Cambridge University Press.

Baayen, R. Harald. 2013. *Package "languageR."* http://cran.r-project.org/web/packages/languageR/languageR.pdf.

Bartoń, Kamil. 2013. *Package "MuMIn."* http://cran.r-project.org/web/packages/MuMIn/index.html.

Bates, Douglas, Martin Maechler & Ben Bolker. 2013. *Package "lme4."* http://cran.r-project.org/web/packages/lme4/lme4.pdf.

Beckman, Mary. 1992. Evidence for Speech Rhythms across Languages. In Yoh'ichi Tokhura, Eric Vatikiotis-Bateson & Yoshinori Sagisaka (eds.), *Speech Perception, Production, and Linguistic Structure*, 457–463. VA: IOS Press.

Beckman, Mary E. & Jan Edwards. 1994. Articulatory evidence for differentiating stress categories. In Patricia A. Keating (ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, 7–33. Cambridge, UK: Cambridge University Press.

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Daniel Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60. 92–111.

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113(2). 1001–1024.

Benor, Sarah Bunin & Roger Levy. 2006. The Chicken or the Egg? A Probabilistic Analysis of English Binomials. *Language* 82(2). 233–278.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow, UK: Longman.

Bloothooft, Gerrit & Loek Groot. 2008. Name clustering on the basis of parental preferences. *Names* 56(3). 111–163.

Bloothooft, Gerrit & David Onland. 2011. Socioeconomic determinants of first names. *Names* 59(1). 25–41.

Bock, Kathryn. 1987. An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language* 26(2). 119–137.

Bock, Kathryn & Willem Levelt. 1994. Language Production: Grammatical Encoding. In M.A. Gernsbacher (ed.), *Handbook of Psycholinguistics*, 741–779. New York, NY: Academic Press, Inc.

Bod, Rens, Jennifer Hay & Stefanie Jannedy (eds.). 2003. *Probabilistic Linguistics*. Cambridge, MA: The MIT Press.

Boers, Frank & Seth Lindstromberg. 2005. Finding ways to make phrase-learning feasible: The mnemonic effect of alliteration. *System* 33(2). 225–238.

Booij, Geert. 1997. Non-derivational phonology meets lexical phonology. In Iggy Roca (ed.), *Derivations and Constraints in Phonology*, 261–288. Oxford: Oxford University Press.

Booij, Geert. 1998. Phonological output constraints in morphology. In Wolfgang Kehrein & Richard Wiese (eds.), *Phonology and Morphology of the Germanic Languages*, 143–163. Tübingen, Germany: Max Niemeyer.

Bowes, Ron. 2010. Facebook username list. http://www.skullsecurity.org/blog/2010/return-of-the-facebook-snatchers.

Bradley, Dianne Christine. 1978. Computational distinctions of vocabulary type. Cambridge, MA: Massachusetts Institute of Technology Ph.D. dissertation.

Breiman, Leo. 2001. Random Forests. *Machine Learning* 45(1). 5–32.

Bresnan, Joan, Jean Caletta, Richard Crouch, Malvina Nissim, Mark Steedman, Tom Wasow & Annie Zaenen. 2002. Paraphrase analysis for improved generation. *LINK project. HCRC Edinburgh*. Stanford, CA: CSLI.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald Baayen. 2007. Predicting the dative alternation. In G Boume, I Kraemer & J Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan & Marilyn Ford. 2010. Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English. *Language* 86(1). 186–213.

Burnham, Kenneth P. & David R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd edition. New York, NY: Springer-Verlag.

Burnham, Kenneth P. & David R. Anderson. 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research* 33(2). 261–304.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Cann, Ronnie. 2000. Functional versus lexical: A cognitive dichotomy. In Robert D. Borsley (ed.), *The Nature and Function of Syntactic Categories*, 37–78. (Syntax and Semantics 32). San Diego, CA: Academic Press.

Carstairs, Andrew. 1990. Phonologically conditioned suppletion. In Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.), *Contemporary Morphology*, 17–23. Walter de Gruyter.

Carstairs-McCarthy, Andrew. 1998. Phonological constraints on morphological rules. In Andrew Spencer & Arnold Zwicky (eds.), *The Handbook of Morphology*, 144–148. Oxford, UK: Blackwell Publishing.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. Harper & Row.

Coetzee, Andries W. & Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language & Linguistic Theory* 26(2). 289–337.

Cohen-Goldberg, Ariel M., Joana Cholin, Michele Miozzo & Brenda Rapp. 2013. The interface between morphology and phonology: Exploring a morpho-phonological deficit in spoken production. *Cognition* 127. 270–286.

Couper-Kuhlen, Elizabeth. 1993. *English speech rhythm: Form and function in everyday verbal interaction*. Amsterdam, The Netherlands: Benjamins.

Cutler, Anne. 1993. Phonological cues to open- and closed-class words in the processing of spoken sentences. *Journal of Psycholinguistic Research* 22(2). 109–131.

Dauer, Rebecca M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11(1). 51–62.

Ehret, Katherina Luisa. 2011. The role of rhythm as a factor in historical genitive variability. Freiburg, Germany: University of Freiburg MA thesis.

Elenbaas, Nine & René Kager. 1999. Ternary Rhythm and the Lapse Constraint. *Phonology* 16(3). 273–329.

Embick, David. 2007. Blocking Effects and Analytic/synthetic Alternations. *Natural Language & Linguistic Theory* 25(1). 1–37.

Embick, David. 2010. *Localism versus Globalism in Morphology and Phonology*. (Linguistic Inquiry Monographs 60). Cambridge, MA: The MIT Press.

Embick, David & Alec Marantz. 2008. Architecture and Blocking. *Linguistic Inquiry* 39(1). 1–53.

English, Leo. 1986. *Tagalog-English dictionary*. Manila: Congregation of the Most Holy Redeemer; distributed by Philippine National Book Store.

Ferreira, Victor S. & L. Robert Slevc. 2007. Grammatical encoding. In Mark Gareth Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*, 453–469. Oxford, UK: Oxford University Press.

Fitt, Susan. 2001. Unisyn lexicon. Centre for Speech Technology Research, University of Edinburgh. www.cstr.ed.ac.uk/projects/unisyn.

Fitzgerald, Colleen M. 2007. An optimality treatment of syntactic inversions in English verse. *Language Sciences* 29. 203–217.

Fougeron, Cécile & Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101(6). 3728–3740.

Frank, Austin & T. Florian Jaeger. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. *Cogsci*. Washington, DC: CogSci.

Fry, Dennis. 1955. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 35. 765–69.

Gahl, Susanne, Yao Yao & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66. 789–806.

Garretson, Gregory, M. Catherine O'Connor, Barbora Skarabela & Marjorie Hogan. 2004. Coding practices used in the project Optimality Typology of Determiner Phrases. http://npcorpus.edu/documentation/index.html.

Garrett, Merrill F. 2000. Remarks on the architecture of language processing systems. In Yosef Grodzinsky, Lewis P. Shapiro & David Swinney (eds.), *Language and the Brain: Representation and Processing*, 3–69. (Foundations in Neuropsychology). San Diego, CA: Academic Press.

Gelman, Andrew. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* 27. 2865–2873.

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz & W. O'Neil (eds.), *Image, Language, Brain*, 95–126. Cambridge, MA: MIT Press.

Gibson, Edward & Evelina Fedorenko. 2010. The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes* 28(1-2). 88–124.

Godfrey, J. Holliman & J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In Y. Miyashita, A. Marantz & W. O'Neil (eds.), *Proceedings of ICASSP-92*, 517–520.

Goldsmith, John. 1976. Autosegmental phonology. Cambridge, MA: Massachusetts Institute of Technology Ph.D. dissertation.

Golston, Chris. 1995. Syntax Outranks Phonology: Evidence from Ancient Greek. *Phonology* 12(3). 343–368.

Gordon, Barry & Alfonso Caramazza. 1982. Lexical Decision for Open- and Closed-Class Words: Failure to Replicate Differential Frequency Sensitivity. *Brain and Language* 15. 143–160.

Gouskova, Maria. 2003. Deriving Economy: Syncope in Optimality Theory. University of Massachusetts, Amherst.

Grabe, Esther & Ee Ling Low. 2002. Acoustic correlates of rhythm class. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7*, vol. 1, 515–546. Berlin, Germany: Mouton de Gruyter.

Grabe, Esther & Paul Warren. 1995. Stress shift: do speakers do it or do listeners hear it? In Bruce Connell & Amalia Arvaniti (eds.), *Phonology and phonetic evidence:*

Papers in Laboratory Phonology IV, 95–110. Cambridge, UK: Cambridge University Press.

Grafmiller, Jason. forthcoming. Variation in English genitives across modality and genres. *English Language and Linguistics*.

Grafmiller, Jason & Stephanie S Shih. 2011. New approaches to end weight. Paper. Paper presented at the Variation and Typology: New trends in Syntactic Research, Helsinki, Finland. http://stanford.edu/~stephsus/GrafmillerandShihHelsinki2011.pdf.

Grahn, Jessica A. & James B. Rowe. 2013. Finding and feeling the musical beat: Striatal dissociations between detection and prediction of regularity. *Cerebral Cortex* 23(4). 913–921.

Greven, Sonja & Thomas Kneib. 2010. On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika* 97(4). 773–789.

Gussenhoven, Carlos. 1991. The English rhythm rule as an accent deletion rule. *Phonology* 8. 1–35.

Halpern, A. 1992. Topics in the Syntax and Placement of Clitics. Stanford University.

Hanson, Kristin & Paul Kiparsky. 1996. A Parametric Theory of Poetic Meter. *Language* 72(2). 287–335.

Hanssen, Esther, Arina Banga, Robert Schreuder & Anneke Neijt. 2013. Semantic and prosodic effects of Dutch linking elements. *Morphology* 23. 7–32.

Harrell, Frank E., Jr. 2009. *Package "Design"*. http://cran.r-project.org/web/packages/Design/index.html.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge, UK: Cambridge University Press.

Hay, Jennifer B. & R. Harald Baayen. 2005. Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences* 9(7). 342–348.

Hayes, Bruce. 1984. The Phonology of Rhythm in English. *Linguistic Inquiry* 15(1). 33–74.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago, IL: The University of Chicago Press.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39(3). 379–440.

Hinrichs, Lars & Benedikt Szmrecsányi. 2007. Recent changes in the function and frequency of standard English genitive constructions: a multivariate analysis of tagged corpora. *English Language and Linguistics* 11(3). 437–474.

Hirschberg, Julia. 1993. Pitch accent in context: predicting intonational prominence from text. *Artificial Intelligence* 63. 305–340.

Hopper, Paul J. & Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge, UK: Cambridge University Press.

Hothorn, Torsten, Kurt Hornik, Carolin Strobl & Achim Zeileis. 2013. *Package "party."* http://cran.r-project.org/package=party.

Hyndman, Rob J. & Yanan Fan. 1996. Sample quantiles in statistical packages. *The American Statistician* 50(4). 361–365.

Ingason, Anton Karl & Laurel MacKenzie. 2011. "Heaviness" as evidence for a derive-and-compare grammar. Poster presentation. Paper presented at the 19th Manchester Phonology Meeting, Manchester, UK. http://www.linguist.is/skjol/mfm2011heaviness.pdf.

Inkelas, Sharon. 1990. *Prosodic Constituency in the Lexicon*. (Outstanding Dissertations in Linguistics). New York, NY: Garland Publishing, Inc.

Inkelas, Sharon. 2014. *Interplay of Phonology and Morphology*. (Oxford Surveys in Syntax and Morphology Series). Oxford, UK: Oxford University Press.

Inkelas, Sharon & Stephanie S Shih. 2013. Unstable surface correspondence as the source of local conspiracies. Paper presented at the NELS 2013, University of Connecticut. http://stanford.edu/~stephsus/InkelasShih_NELS_2013_handout.pdf.

Inkelas, Sharon & Draga Zec. 1993. Auxiliary reduction without empty categories: a prosodic account. *Working Papers of the Cornell Phonetics Laboratory* (8), 205–254. Cornell University.

Inkelas, Sharon & Draga Zec. 1995. Syntax-phonology interface. In John A Goldsmith (ed.), *The Handbook of Phonological Theory*, 535–549. Cambridge, Mass., and Oxford, UK: Blackwell.

Itô, Junko & Armin Mester. 2009. The extended prosodic word. In Janet Grijzenhout & Barış Kabak (eds.), *Phonological Domains: Universals and Deviations*, 135–194. Berlin, Germany: Mouton de Gruyter.

Jaeger, T. Florian. 2005. “That” indicates anticipated production difficulty: Evidence from disfluencies. *Proceedings of Disfluency in Spontaneous Speech*, 103–109.

Jaeger, T. Florian. 2006. Phonological optimization and syntactic variation: The case of optional that. *Proceedings of the 32nd Annual Meeting of the Berkeley Linguistic Society*. Berkeley, CA.

Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61. 23–62.

Johnson, Keith. 2004a. Massive reduction in conversational American English. *Spontaneous Speech: Data and Analysis.*, 29–54. Tokyo, Japan: The National International Institute for Japanese Language.

Johnson, Keith. 2004b. Aligning phonetic transcriptions with their citation forms. *Acoustics Research Letters Online* 5. 19–24.

Johnson, Keith & John W. Mullenix (eds.). 1997. *Talker Variability in Speech Processing*. San Diego, CA: Academic Press, Inc.

Jurafsky, Daniel & James H. Martin. 2009. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. 2nd ed. Upper Saddle River, NJ: Pearson Education, Inc.

Kager, René. 1996. On affix allomorphy and syllable counting. In Ursula Kleinhenz (ed.), *Interfaces in phonology*, 155–171. Berlin: Akademie Verlag.

Kaisse, Ellen M. 1985. *Connected Speech: The Interaction of Syntax and Phonology*. Orlando, Florida: Academic Press, Inc.

Kelly, Michael H & David C Rubin. 1988. Natural rhythmic patterns in English verse: Evidence from child counting-out rhymes. *Journal of Memory and Language* 27(6). 718–740.

Kelly, Michael H. & J. Kathryn Bock. 1988. Stress in time. *Journal of Experimental Psychology: Human Perception and Performance* 14(3). 389–403.

Kendall, Tyler, Joan Bresnan & Gerard van Herk. 2011. The Dative Alternation in African American English: Researching Syntactic Variation and Change across Sociolinguistic Datasets. *Corpus Linguistics and Linguistic Theory* 7(2). 229–244.

Kim, Yuni. 2010. Phonological and morphological conditions on affix order in Huave. *Morphology* 20(1). 133–163.

Kiparsky, Paul. 1977. The Rhythmic Structure of English Verse. *Linguistic Inquiry* 8(2). 189–247.

Kiparsky, Paul. 1982. Lexical Morphology and Phonology. In I. -S. Yang (ed.), *Linguistics in the Morning Calm*, vol. 2, 3–91. Seoul: Hanshin Publishing Company.

Kiparsky, Paul. 1985. Some Consequences of Lexical Phonology. *Phonology Yearbook* 2. 85–138.

Kiparsky, Paul. 2005. Blocking and periphrasis in inflectional paradigms. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2004*, 113–135. Amsterdam, Netherlands: Springer.

Kisseberth, Charles. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1. 291–306.

Klein, Thomas B. 2003. Syllable structure and lexical markedness in creole morphophonology: Determiner allomorphy in Haitian and elsewhere. In Ingo Plag (ed.), *Phonology and morphology of creole languages*, vol. 478, 209. Tuebingen, Germany: Max Niemeyer Verlag.

Kreyer, Rolf. 2003. Genitive and of-construction in modern written English: Processability and human involvement. *International Journal of Corpus Linguistics* 8(2). 169–207.

Kullback, Soloman & Richard A. Leibler. 1951. On Information and Sufficiency. *Annals of Mathematical Studies* 2. 79–86.

Kuperman, Victor & Joan Bresnan. 2012. The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language* 66. 588–611.

Labov, William. 1969. Contraction, Deletion, and Inherent Variability of the English Copula. *Language* 45(4). 715–762.

Labov, William. 1972. The reflection of social processes in linguistic structures. *Sociolinguistic Patterns*, 110–121. Philadelphia: University of Pennsylvania Press.

Large, Edward W. & Caroline Palmer. 2002. Perceiving temporal regularity in music. *Cognitive Science* 26. 1–37.

Leben, William R. 1973. Suprasegmental phonology. Cambridge, MA: Massachusetts Institute of Technology Ph.D. dissertation.

Lebreton, Jean-Dominique, Kenneth P. Burnham, Jean Clobert & David R. Anderson. 1992. Modeling Survival and Testing Biological Hypotheses Using Marked Animals: A Unified Approach with Case Studies. *Ecological Monographs* 62(1). 67–118.

Lee, Ming-Wei & Julie Gibbons. 2007. Rhythmic alternation and the optional complementiser in English: New evidence of phonological influence on grammatical encoding. *Cognition* 105. 446–456.

Lehiste, Ilse. 1977. Isochrony revisited. *Journal of Phonetics* 5(3). 253–263.

Lerdahl, Fred & Ray Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, Mass.: MIT Press.

Levelt, Willem. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, Willem J. M. 2001. Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, vol. 98(23), 13464–13471.

Liberman, Mark. 1975. The Intonational System of English. MIT.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8(2). 249–336.

Liu, Baolin, Zhixing Jin, Zhongning Wang & Guangning Wu. 2010. Chinese function words grammaticalized from content words: Evidence from ERPs. *Journal of Neurolinguistics* 23(6). 663–675.

Manning, Christopher D. 2003. Probabilistic syntax. In Rens Bod, Jennifer Hay & Stefanie Jannedy (eds.), *Probabilistic linguistics*, 289–341. Cambridge, MA and London: MIT Press.

Marcus, Mitchell P., Beatrice Santorini & Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19(2). 313–330.

Martin, Andrew T. 2005. The effects of distance on lexical bias: sibilant harmony in Navajo compounds. M.A. thesis. University of California, Los Angeles, ms.

Martin, Andrew T. 2007. The Evolving Lexicon. Los Angeles, CA: University of California, Los Angeles Ph.D. dissertation.

Martin, Andrew T. 2011. Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language* 87(4). 751–770.

Mateos, Pablo & Ken Tucker. 2008. Forenames and surnames in Spain in 2004. *Names* 56(3). 165–184.

McCarthy, John J. 2008. The serial interaction of stress and syncope. *Natural Language & Linguistic Theory* 26(3). 499–546.

McCarthy, John J & Alan Prince. 1993a. Generalized Alignment. In Geert E Booij & Jaap van Marle (eds.), *Yearbook of Morphology*, 79–153. Dordrecht: Kluwer.

McCarthy, John J & Alan Prince. 1993b. *Prosodic morphology I: Constraint interaction and satisfaction*. New Brunswick: Rutgers University Center for Cognitive Science.

McCarthy, John J & Matthew Wolf. 2005. Less than zero: Correspondence and the Null Output. Amherst, MA.

McDonald, Janet L., Kathryn Bock & Michael H. Kelly. 1993. Word and World Order: Semantic, Phonological, and Metrical Determinants of Serial Position. *Cognitive Psychology* 25. 188–230.

McPherson, Laura & Bruce Hayes. 2013. Relating application frequency to morphological structure: the case of Tommo So vowel harmony. University of California, Los Angeles, ms. http://www.linguistics.ucla.edu/people/grads/McPherson/Laura_McPhersons_Homepage/Downloadable_papers_files/TommoSoVowelHarmonySept2013.pdf.

Menn, Lise & Brian MacWhinney. 1984. The Repeated Morph Constraint: Toward an Explanation. *Language* 60(3). 519–541.

Mester, R. Armin. 1994. The quantitative trochee in Latin. *Natural Language & Linguistic Theory* 12(1). 1–61.

Mielke, Jeff. 2012. A phonetically based metric of sound similarity. *Lingua* 122(2). 145–163.

Miller, Philip H, Geoffrey K Pullum & Arnold M Zwicky. 1997. The Principle of Phonology-Free Syntax: Four Apparent Counterexamples in French. *Journal of Linguistics* 33(1). 67–90.

Mollin, Sandra. 2012. Revisiting binomial order in English: ordering constraints and reversibility. *English Language and Linguistics* 16(1). 81–103.

Mortensen, David. 2006. Logical and substantive scales in phonology. Berkeley, CA: University of California, Berkeley Ph.D. dissertation.

Nenkova, Ani, Jason M. Brenier, Anubha Kothari, Sasha Calhoun, Laura Whitton, David Beaver & Daniel Jurafsky. 2007. To memorize or to predict: Prominence labeling in conversational speech. *Proceedings of NAACL HLT 2007*, 9–16. Rochester, NY.

Nespor, Marina & Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris.

Nespor, Marina & Irene Vogel. 1989. On Clashes and Lapses. *Phonology* 6(1). 69–116.

Nevins, Andrew. 2011. Phonologically-conditioned allomorph selection. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, vol. IV, 2357–2382. Malden, MA and Oxford, UK: Blackwell Publishing.

Odden, David. 1993. Interaction between modules in Lexical Phonology. In Sharon Hargus & Ellen Kaisse (eds.), *Phonetics and Phonology 4: Studies in Lexical Phonology*, 111–144. San Diego, CA: Academic Press, Inc.

Orgun, Cemil Orhan & Ronald L Sprouse. 1999. From MParse to Control: Deriving Ungrammaticality. *Phonology* 16(2). 191–224.

Osselton, Noel E. 1988. Thematic genitives. In G. Nixon & J. Honey (eds.), *An Historic Tongue: Studies in English Linguistics in Memory of Barbara Strang*. London, UK: Routledge.

Ostendorf, Mari, Izhak Shafran, Stefanie Shattuck-Hufnagel, Lesley Carmichael & William Byrne. 2001. A prosodically labeled database of spontaneous speech. *ISCA Tutorial and Research Workshop on Prosody and Speech Recognition and Understanding*. Red Bank, NJ: International Speech Communication Association. http://www.isca-speech.org/archive_open/archive_papers/prosody_2001/prsr_022.pdf.

Palmer, Caroline & Michael H Kelly. 1992. Linguistic prosody and musical meter in song. *Journal of Memory and Language* 31(4). 525–542.

Paster, Mary. 2006. Phonological conditions on affixation. Berkeley, CA: University of California, Berkeley Ph.D. dissertation.

Paster, Mary. 2009a. Phonologically-Conditioned Affix Order as an Illusory Phenomenon. Paper. Paper presented at the Workshop on the Division of Labour between Morphology and Phonology, Amsterdam, Netherlands. http://www.uni-leipzig.de/~exponet/Slides/Amsterdam/Paster.pdf.

Paster, Mary. 2009b. Explaining phonological conditions on affixation: Evidence from suppletive allomorphy and affix ordering. *Word Structure* 2(1). 18–37.

Pater, Joe. 1999. Austronesian nasal substitution and other NC effects. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.), *The Prosody-Morphology Interface*, 310–343. Cambridge: Cambridge University Press.

Pike, Kenneth L. 1945. *The Intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Pitt, Mark A., Laura Dilley, Keith Johnson, Scott Kiesling, William D. Raymond, Elizabeth Hume & Eric Fosler-Lussier. 2007. Buckeye Corpus of Conversational Speech (2nd release). Department of Psychology, Ohio State University. www.buckeyecorpus.osu.edu.

Plag, Ingo. 1999. *Morphological productivity: structural constraints in English derivation*. Berlin: Mouton de Gruyter.

Poser, William J. 1992. Blocking of Phrasal Constructions by Lexical Items. In Ivan A. Sag & Anna Szabolcsi (eds.), *Lexical Matters*, 111–130. Stanford, CA: CSLI.

Prince, Alan S. 1983. Relating to the Grid. *Linguistic Inquiry* 14(1). 19–100.

Prince, Alan & Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in generative grammar*. Malden, Mass., and Oxford, UK: Blackwell.

Pullum, Geoffrey K. 2011. A warm welcome to a new determinative. *Language Log*. http://languagelog.ldc.upenn.edu/nll/?p=3385.

Pullum, Geoffrey & Arnold Zwicky. 1988. The syntax-phonology interface. In F. Newmeyer (ed.), *Linguistics: The Cambridge Survey*, vol. 1, 255–280. Cambridge: Cambridge University Press.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.

R Development Core Team. 2010. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. www.R-project.org.

Raffelsiefen, Renate. 1996. Gaps in word formation. In Ursula Kleinhenz (ed.), *Interfaces in phonology*, 194–209. Berlin: Akademie Verlag.

Raffelsiefen, Renate. 1999. Phonological constraints on English word formation. In Geert E Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1998*, 225–287. (Yearbook of Morphology 8). Springer.

Raffelsiefen, Renate. 2004. Absolute ill-formedness and other morphophonological effects. *Phonology* 21(1). 91–142.

Ramscar, Michael, Asha Smith, Richard Futrell, Peter Hendrix, Melody Dye, Harald Baayen & Rebecca Starr. 2013. The “universal” structure of name grammars and the impact of social engineering on the evolution of natural information systems. *Cogsci 2013*. http://mindmodeling.org/cogsci2013/papers/0578/paper0578.pdf.

Ramus, Frank, Marina Nespor & Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73. 265–292.

Raymond, William D. 2003. An analysis of coding consistency in the transcription of spontaneous speech from the Buckeye corpus. *Proceedings of the Workshop on Spontaneous Speech: Data and Analysis*. Tokyo, Japan. http://vic.ling.ohio-state.edu/SSDA.ms.pdf.

Repp, Bruno H. 2010. Do metrical accents create illusory phenomenal accents? *Attention, Perception, and Psychophysics* 72(5). 265–292.

Rice, Keren. 2011. Principles of affix ordering: An overview. *Word Structure* 4(2). 169–200.

Rosenbach, Anette. 2002. *Genitive variation in English: conceptual factors in synchronic and diachronic studies*. Berlin, New York: Mouton de Gruyter`.

Rosenbach, Anette. 2005. Animacy versus Weight as Determinants of Grammatical Variation in English. *Language* 81(3). 613–644.

Rosenbach, Anette. 2007. Emerging variation: determiner genitives and noun modifiers in English. *English Language and Linguistics* 11(1). 143–189.

Rosenbach, Anette. 2008. Animacy and grammatical variation—Findings from English genitive variation. *Lingua* 118(2). 151–171.

Rounds, Carol. 2001. *Hungarian: An Essential Grammar*. New York, NY: Routledge.

Ryan, Kevin M. 2013. Onset weight, word weight, and the perceptual interval. Paper presentation. Paper presented at the Phonology 2013, University of Massachusetts, Amherst.

Sag, Ivan A. & Thomas Wasow. 2011. Performance-compatible competence grammar. In Robert D. Borsley & Kersti Börjars (eds.), *Non-Transformational Syntax: Formal and Explicit Models of Grammar*, 359–377. Malden, MA and Oxford, UK: Wiley-Blackwell.

Santorini, Beatrice. 1990. *Part-of-Speech Tagging Guidelines for the Penn Treebank Project (3rd Revision)*. (Department of Computer & Information Science Technical Reports (CIS)). Philadelphia, PA: University of Pennsylvania. http://repository.upenn.edu/cis_reports/570.

Schachter, Paul & Fe T. Otanes. 1972. *Tagalog Reference Grammar*. Berkeley and Los Angeles, CA: University of California Press. http://www.ucpress.edu/op.php?isbn=9780520049437.

Schlüter, Julia. 2005. *Rhythmic Grammar: The Influence of Rhythm on Grammatical Variation and Change in English*. Berlin, Germany: Mouton de Gruyter.

Schutze, Carson. 1994. Serbo-croatian second position clitic placement and the phonology-syntax interface. *MIT Working Papers in Linguistics* 21. 373–473.

Segalowitz, Sidney J. & Korri C. Lane. 2000. Lexical access of function versus content words. *Brain and Language* 75. 376–389.

Segui, Juan, Jacques Mehler, Uli Frauenfelder & John Morton. 1982. The word frequency effect and lexical access. *Neuropsychologia* 20(6). 615–627.

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology* 3. 371–405.

Selkirk, Elisabeth O. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: The MIT Press.

Selkirk, Elisabeth O. 1996. The prosodic structure of function words. In J. Martin & K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, 187–213. Mahwah, NJ: Lawrence Erlbaum.

Shih, Stephanie S, Jason Grafmiller, Richard Futrell & Joan Bresnan. to appear. Rhythm's role in predicting genitive alternation choice in spoken English. In Ralf Vogel & Ruben van de Vijver (eds.), *Rhythm in phonetics, grammar, and cognition*.

Smith, Brian W. 2013. -(a)licious: a new morpheme obeys existing constraints. Ms. University of Massachusetts, Amherst, ms. http://people.umass.edu/bwsmith/papers/bws_licious.pdf.

Stampe, David. 1973. A Dissertation on Natural Phonology. University of Chicago.

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9. 307–317.

Strobl, Carolin, Torsten Hothorn & Achim Zeileis. 2009. Party on! *The R Journal* 1(2). 14–17.

Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4). 323–348.

Sumner, Meghan. 2011. The role of variation in the perception of accented speech. *Cognition* 119(1). 131–136.

Sumner, Meghan & Reiko Kataoka. 2013. Effects of phonetically-cued talker variation on semantic encoding. *The Journal of the Acoustical Society of America* 134(6). EL485–EL491.

Svartvik, Jan & Randolph Quirk (eds.). 1980. *A corpus of English conversation*. (Lund Studies in English 56). Lund: Lund University Press.

Sweet, Henry. 1876. Words, Logic, and Grammar. In H.C. Wyld (ed.), *The Collected Papers of Henry Sweet*. London, UK: Oxford University Press.

Szmrecsányi, Benedikt. 2004. On Operationalizing Syntactic Complexity. *7es Journées internationales d'Analyse statistique des Données Textuelles*, 1031–1038. http://www.benszm.net/omnibuslit/Szmrecsanyi2004.pdf.

Szmrecsányi, Benedikt. 2006. *Morphosyntactic persistence in spoken English: A corpus study of the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin and New York: Mouton de Gruyter.

Szmrecsányi, Benedikt & Lars Hinrichs. 2008. Probabilistic determinants of genitive variation in spoken and written English: A multivariate comparison across time, space, and genres. In Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta & Minna Korhonen (eds.), *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*. Amsterdam, Netherlands: Benjamins.

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24. 135–178.

Tagliamonte, Sali & Lidia Jarmasz. 2008. Variation and change in the English genitive: A sociolinguistic perspective. Paper. Paper presented at the Linguistic Society of America Annual Meeting, Chicago, IL.

Taylor, John R. 1996. *Possessives in English*. Oxford, UK: Clarendon Press.

Temperley, David. 2006. Minimization of dependency length in written English. *Cognition* 105. 300–333.

Temperley, David. 2009. Distributional Stress Regularity: A Corpus Study. *Journal of Psycholinguistic Research* 38(1). 75–92.

Tilsen, Sam. 2011. Metrical regularity facilitates speech planning and production. *Laboratory Phonology* 2(1). 185–218.

Tilsen, Sam & Amalia Arvaniti. 2013. Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of the Acoustical Society of America* 134(1). 628–639.

Tilsen, Sam, Emily Cibelli, Shinae Kang, John Houde & Srikantan Nagarajan. 2012. Acoustic analysis of the effects of metrical regularity on interval durations. *Cornell Working Papers in Phonetics and Phonology*. http://conf.ling.cornell.edu/~tilsen/papers/Tilsen%20et%20al.%20-%202012%20-%20effects%20of%20metrical%20regularity.pdf.

Tilsen, Sam & Keith Johnson. 2008. Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America* 124. EL34–39.

Tomlinson, John M., Jr., Qiang Liu & Jean E. Fox Tree. 2013. The perceptual nature of stress shifts. *Language and Cognitive Processes*. 1–13.

Tucker, D.K. 2001. Distribution of Forenames, Surnames, and Forename-Surname Pairs in the United States. *Names* 49(2). 69–96.

Uhmann, Susanne. 1996. On rhythm in everyday German conversation: Beat clashes in assessment utterances. *Studies in Interactional Sociolinguistics* 12. 303–365.

Ussishkin, Adam & Andrew Wedel. 2002. Neighborhood Density and the Root-Affix Distinction. In Masako Hirotani (ed.), *Proceedings of NELS 32*.

Ussishkin, Adam & Andrew Wedel. 2009. Lexical access, effective contrast, and patterns in the lexicon. In Paul Boersma & Silke Hamann (eds.), *Phonology in Perception*, vol. 15, 267–292. (Phonology and Phonetics). Berlin, Germany: Walter de Gruyter.

Vaida, Florin & Suzette Blanchard. 2005. Conditional Akaike information for mixed-effects models. *Biometrika* 92(2). 351–370.

Vogel, Irene & Istvan Kenesei. 1990. Syntax and semantics in phonology. In Sharon Inkelas & Draga Zec (eds.), *The Phonology-Syntax Connection*, 339–364. Chicago: University of Chicago Press.

Walter, Mary & T. Jaeger. 2005. Constraints on Optional *that:* A Strong Word Form OCP Effect. *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 41(1). 505–519.

Ward, Jr., Joe H. 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 58(301). 236–244.

Wasow, Thomas. 2009. Gradient data and gradient grammars. *Proceedings of the 43rd Annual Meeting of the Chicago Linguistics Society*, 255–271. Chicago, IL.

Wasow, Thomas, Rebecca Greene & Roger Levy. 2012. Optional *to* and Prosody. Poster presentation. Paper presented at the CUNY, Stanford, CA.

Wasow, Tom. 2002. *Postverbal Behavior*. Stanford, CA: CSLI Publications.

Weide, Robert. 1993. *Carnegie Mellon Pronouncing Dictionary [cmudict.0.2]*.

Whissell, Cynthia. 2001. Sound and emotion in given names. *Names* 49(2). 97–120.

Wilen, Joan & Lydia Wilen. 1993. *The Perfect Name for the Perfect Baby*. Ballatine Books.

Wolf, Matthew. 2008. Optimal Interleaving: serial phonology-morphology interaction in a constraint-based model. University of Massachusetts Amherst.

Wright, Saundra K., Jennifer Hay & Tessa Bent. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics* 43(3). 531–561.

Yao, Yao. 2011. The Effects of Phonological Neighborhoods on Pronunciation Variation in Conversational Speech. Berkeley, CA: University of California, Berkeley Ph.D. dissertation.

Yip, Moira. 1998. Identity avoidance in phonology and morphology. In Steven G Lapointe, Diane K Brentari & Patrick M Farrell (eds.), *Morphology and its Relation to Phonology and Syntax*, 216–246. Stanford, CA: CSLI Publications.

Youmans, Gilbert. 1983. Generative Tests for Generative Meter. *Language* 59(1). 67–92.

Yuan, Jiahong, Jason M. Brenier & Daniel Jurafsky. 2005. Pitch Accent Prediction: Effects of Genre and Speaker. *Proceedings of Interspeech*.

Zaenen, Annie, Jean Carlette, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M. Catherine O'Connor & Tom Wasow. 2004. Animacy encoding in English: Why and how. In D. Byron & B. Webber (eds.), *2004 ACI workshop on discourse annotation*, 118–125. Barcelona, Spain.

Zec, Draga & Dušica Filipović-Đurđević. to appear. The Role of Prosody in Clitic Placement. In Vera Gribanova & Stephanie S Shih (eds.), *The morphosyntax-phonology connection: locality and directionality at the interface*. Oxford, UK: Oxford University Press.

Zec, Draga & Sharon Inkelas. 1990. Prosodically constrained syntax. In Sharon Inkelas & Draga Zec (eds.), *The Phonology-Syntax Connection*, 365–378. Chicago: University of Chicago Press.

Zipf, George Kingsley. 1929. Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology* 40. 1–95.

Zubizarreta, Maria Luisa. 1998. *Prosody, Focus, and Word Order*. (Linguistic Inquiry Monographs 33). Cambridge, MA: MIT Press.

Zuraw, Kie. 2000. Patterned exceptions in phonology. University of California, Los Angeles PhD dissertation.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language & Linguistic Theory* 28(2). 417–472.

Zuraw, Kie & Yu-An Lu. 2009. Diverse repairs for multiple labial consonants. *Natural Language and Linguistic Theory* 27. 197–224.

Zwicky, Arnold M. 1987. Suppressing the Z's. *Journal of Linguistics* 23. 133–148.

Zwicky, Arnold M. & Geoffrey Pullum. 1986a. The principle of phonology-free syntax: introductory remarks. *Interfaces*, vol. 32, 63–91. (Ohio State University Working Papers in Linguistics).

Zwicky, Arnold M. & Geoffrey Pullum. 1986b. Two spurious counterexamples to the Principle of Phonology-Free Syntax. *Interfaces*, vol. 32, 92–99. (Ohio State University Working Papers in Linguistics).